

Beyond Infinity ∞

An expedition to the outer-limits of the mathematical universe

超越无穷大

跨越思维极限数学之旅

尤金妮娅·程（Eugenia Cheng）——著



中信出版集团

版权信息

书名:超越无穷大：一次跨越数学边界的冒险之旅

作者:[英]尤金妮娅·程

译者:杜娟

ISBN:9787508685663

中信出版集团制作发行

题记

纪念萨拉·巴德尔（Sara Al-Bader）

她教会我有限的生命里可以容纳无穷的爱

序言

我讨厌机场。

我觉得机场让人特别有压力，拥挤而且吵闹。机场里总是充斥着过多的人、过多的队列、过少的作为和无处不在诱惑着我的不健康的食物。非常不幸的是，旅程总是从机场开始，这让我有一点儿害怕旅行。旅行应该是一个令人兴奋的有关发现的过程，坐着飞机去一个新的地方，应该是一段壮美而充满魔力的经历。但是机场和狭窄的经济舱座椅经常会毁掉这一切。

学习数学应该也算得上一段令人兴奋的发现之旅，同样壮美且充满魔力，尽管它的开端也经常会毁掉一切。因为一开始你总是会遇见大量的事实、公式、让人感到压力的测试和等待解决的乏味问题。

相较坐飞机，我喜欢乘船旅行。

我喜欢身处开阔的水面，感受风吹过我的脸庞，看着远处的楼群和海岸线，但是又不用靠近。我喜欢一直朝着地平线前进，但是又永远不会到达。我喜欢感受自然的力量，但是又不用完全受大自然的摆布。我不是一个水手，所以通常掌控船的另有其人。偶尔我会遇到一些能够操作的船，这样一来，能够发挥我自己的能力就成了一种奖励。我曾经划着一艘小的手划艇沿着包围着一座小小的法国城堡的护城河漫游，我也曾经沿着阿姆斯特丹的运河踩脚踏船，我还曾经沿着康河撑篙。不过，在一次失足落水之后，我就再也不到康河撑篙了，这和有些人最初在数学领域经历了一些挫折之后就再也不碰数学是一样的。我曾经在悉尼和洛杉矶乘船去看生活在离海岸很远的地方的大鲸鱼，也曾经在威尔士乘船去离海岸很远的地方看海豹和其他野生生物。我小的时候总会和家人一起坐着渡轮跨过英吉利海峡到法国度假，直到几乎不可能建成的欧洲之星变成现实。由此可见，我们人类是多么容易把之前看来几乎不可能的事情当作理所应当！

现在，我很少再为了到达某个目的地而乘船了。相反，我的目的就是享受乘船的过程，欣赏沿海风光和大自然，偶尔发挥一下我自己的能动性。一个例外就是泰晤士河上的渡轮，因为乘坐泰晤士河上的渡轮是伦敦市中心一种令人非常享受的通勤方式。它既能让乘客享受到乘船的乐趣，又能帮助乘客到达目的地。

在某种程度上，我喜欢抽象数学这件事和我喜欢乘船有点儿类似。对于我来说，这两者都超越了到达一个目的地的范畴，更多的是乐趣、锻炼头

脑、与数学交流和欣赏数学之美。这本书是一个通往神秘而壮美的“无穷”世界的旅程。我们即将看到的风景会让我们大开眼界，惊叹不已，甚至有的时候会让我们觉得不可思议。我们将会沉浸在数学的魔力里，但是又不用完全受其摆布。我们将会朝着人类思想的地平线前进，但是又永远不会到达。

第一部分　旅程

1　数学世界里的尼斯湖水怪

无穷就像尼斯湖水怪，以其令人惊叹的体型和难以捉摸的个性，吸引人们展开想象。无穷是一场梦，一个巨大的由无穷无尽的时间和空间所构成的迷幻世界。无穷是一个黑暗森林，在里面，你会遇见超越想象的生物、纠缠在一起的灌木丛和突然照射进来的阳光。无穷是一个环形，它在我们面前呈现为一个无穷无尽的螺旋。

我们的生活是有限的，我们的头脑是有限的，我们所处其中的世界也是有限的。但是我们仍旧能够瞥见我们周围的无穷。我小时候生活的房子中间有一个火炉，火炉上有一个烟囱。所有的房间都围绕着这个火炉连接在一起。这意味着我和妹妹可以一圈一圈地相互追逐，没有终点，感觉上就好像我们生活在一个无穷大的房子里一样。环形让人们可以在有限的空间里开启一个无穷的旅程。这个原理不仅仅被用在了孩子们的相互追逐上，还被用在了汽车赛道上和粒子对撞机上。

后来，我的母亲教我使用频谱计算机编程。直到现在，每当想起我最喜欢的计算机小程序时，我都还是会不由自主地笑出来。

10 打印输出“你好”

20 返回到10

这些计算机语句能够产生一个无穷无尽的循环。当然，这是一个抽象概念的循环，而不是一个物理上的循环。每到这种时候，我就会点击“开始”，然后非常兴奋地看着“你好”这个词在屏幕上滚动。因为我知道，除非我点击“结束”，否则这个过程就会永远重复下去。我是那种不会轻易感到厌烦的孩子。我每天都会这么做，而不觉得自己应该赶紧写一些更加有用的程序。不幸的是，这也使得我的编程能力从来没有真正进步过。无穷的耐心导致了一个奇怪的后果。

我的这个简短但能产生冗长的打印输出结果的抽象循环背后的原理是：程序会自己回到原点。自我索引让我们能够从另一个角度窥测无穷。分形是用和自身形状相同的形状构建的形状。当你把其中的一部分放大的时候，你看到的将是相同的形状。为了达到这一目的，这个形状的细节需要能够“永远”保持下去。毋庸置疑，这些细节一定会超越我们能够描绘和能够看到的极限。图1–1描绘了一些分形树和著名的谢尔宾斯基三角（Sierpinski triangle）的最初几级。

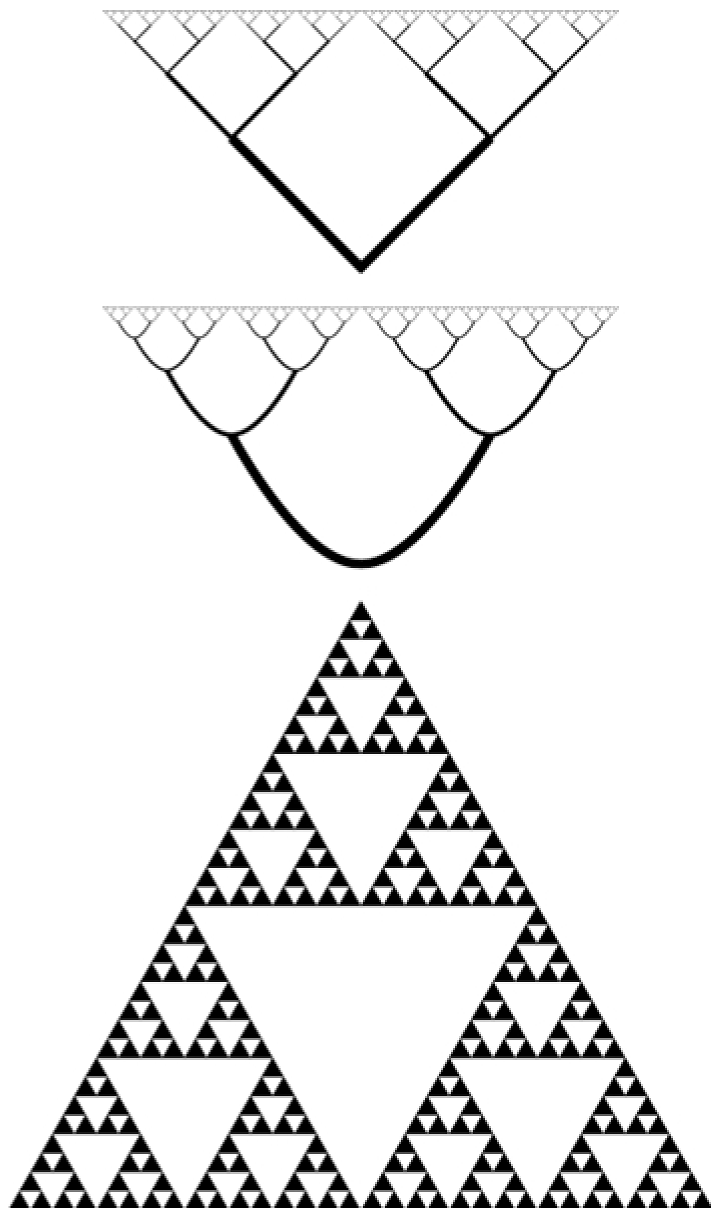


图1–1

如果你把两面镜子对在一起，你看见的将不再仅仅是自己的镜像，还有你的镜像的镜像。你的镜像的镜像又会产生出自己的镜像。只要你调整镜子的角度，这些镜像就会一直叠加下去。每一个镜像都会比前面的镜像小，

从理论上讲，它们会像分形一样“无穷”延续下去。

我们能够从环形和自我索引中看到无穷，同样也能从镜子里的镜像越来越小这件事情上看到无穷。孩子们为了能够让自己一直吃到蛋糕，每次都只吃剩余的蛋糕的一半。一群人在分享一个蛋糕的时候，都很谦让，不愿意吃最后的一块，所以每个人都吃剩余部分的一半。有人告诉我日语里有一个专门形容这个现象的词——“enryo no katamari”，意思就是大家都很谦让而不愿意吃的最后一点蛋糕。

我们并不确切地知道宇宙是不是无穷的。但是我喜欢抬头看教堂的尖塔，让自己相信塔尖的两侧是平行的，而这个塔无穷地冲往天际，直至无穷。我们的生活是有限的，但是不朽的传奇和神话故事能够穿越时间、跨越文化。

对于无穷，我们有如此多需要探索的，就像是尼斯湖波光粼粼的水面下可能藏着也可能没藏着的那个巨大的、古老的、神秘的怪物。这个被我们称为无穷的怪物到底是什么？当我们说“永远”这个听起来平淡无奇的词的时候，我们所指的到底是什么？在我们所生活的这个缺乏耐心的世界里，人们总是很夸张地使用“永远”这个词。可能我们刚在网络上等待了两分钟就会说，“我看起来要永远在线等待了！”如果网页三秒钟没有打开，我们就会说，“这个网页难道永远也打不开吗？”西班牙巴斯克的作家阿玛妮亚·家本桃修（Amaia Gabantxo）告诉我，在巴斯克语里，11对应的词是“hamaika”，这个词同时也有无穷的意思。我的一个来自巴斯克的朋友证实了这一点。他观察了一个橱柜，然后宣称里面有“4瓶2013年的、10瓶2014年的和许多瓶2015年的”自制果酱。很明显，超过10就会被认为是无穷了。我的研究领域是高维范畴理论。在这里，“高”通常意味着三维或者更多，其中也包含无穷。换句话说，从三到无穷都被囊括在了同一个字里。

我们在自己的平凡生活里思考的无穷可能是很梦幻且令人兴奋的，但是一经仔细研究，这些无穷就消失了。就好像彩虹一样，无论你怎样努力也没办法触摸到它。而且，这些无穷往往会造成矛盾、冲突、不可逾越的鸿沟和黑暗的陷阱。在后面的讨论中我们会看到，它们其实禁不起严格的逻辑检验。

数学的一个作用就是解释我们周围世界中出现的各种现象，特别是会在很多不同的地方发生的同样的现象。如果同一个观点能够联系很多不同的情形，数学家就会冲进来试图寻找能将这些情形统一起来的大一统的理论。一旦他们找到了这个理论，我们就能更好地理解这些情形背后的相同点。无穷就是这样的一个观点。它作为一个引人遐想的观点无处不在。它看起来好像也能像其他的数学观点一样最终被统一起来，就好像长度、体积或

者数量一样。但是，为什么当我们把这些简单的数学概念延伸到无穷的时候就这么困难呢？这正是这本书的主旨：为什么这件事这么困难？数学家为此付出的努力最终带来了什么？在这个探索无穷的旅程中我们能看到什么？

无穷的本能

无穷是一个很容易想象但是很难具体描述的概念。小孩子能够很快地理解无穷这个概念，而数学家们却花了几千年的时间从纯技术的层面遵循严格的逻辑论证无穷。下面是一些可能会让我们联想到无穷的事物。孩子们经常会产生这些关于无穷的联想：

无穷会永远持续下去。

无穷比最大的数字还要大。

无穷比我们能想到的任何巨大的事物都更大。

无穷加一，它还是无穷。

无穷加无穷，它还是无穷。

无穷乘以无穷，它依旧是无穷。

孩子们第一次接触无穷这个概念的时候可能会非常激动。他们学习从一数到十，再数到二十，然后学着数到一百、一千、一万、一亿。如果你问一个小孩最大的数字是多少，他们通常会说"一亿"。如果你接着问一亿零一是不是更大呢，他们的眼睛通常会因诧异而瞪大。

无论他们想到的是多大的数字，你都可以加上一然后获得一个更大的数字。说服他们接受这一点并不困难。这说明了一个道理，那就是，最大的数字是不存在的。数字可以一直加下去！但是，一共有多少个数字呢？无穷这个概念由此生成了。

也许一些孩子最初接触无穷这个概念是在电影《玩具总动员》中。他们听见巴斯光年说，"飞向无穷浩瀚的宇宙！"（"To infinity…and beyond!"）这个口号听着就很激动人心。但是，在我还是个孩子的时候，《玩具总动员》这个电影还没有制作出来。我是通过前面描述的循环、家里的物理回路和我最喜欢的计算机程序中的抽象循环来理解无穷的。

一旦孩子们开始思考无穷，他们就会提出很多非常难以回答的关于无穷的问题。无穷是什么？是一个数字吗？是一个地点吗？如果不是地点的话，

我们怎么能像巴斯光年说的那样飞向无穷呢？

从孩子们在学校里听见无穷这个概念起，问题就开始了。一除以零是不是等于无穷？一除以无穷是不是等于零？如果无穷加上一还是无穷的话，那么无穷减去无穷是什么呢？

面对孩子们提出的这些看起来无法回答的数学问题，大人们可能会觉得难为情。因为大人们总会觉得自己需要知道所有的答案。但是数学教育家和创新者克里斯托弗·丹尼尔森说，学习的一个重要的方面就是能够提出新的问题，这比陈述新的事实更加重要。在数学领域，总会有更多的问题。即便是数学非常好的人或者在大学学习数学专业的人，甚至每天从事数学研究的数学家，也总会发现更多的尚且没有得到解答的关于无穷的问题。

无穷的不可思议

下面是一些我最喜欢的令人费解的题目或结论。我们将会在后面的章节中探讨这些题目。

- 如果你有一个有无穷房间的旅馆，而这个旅馆已经客满了，你是否还可以通过将每一个客人挪后一个房间来容纳下一个客人？

- 如果一个彩票机里面有无穷多个号码球，你中奖的概率有多大？

- 一些无穷比其他的无穷要大！

- 无穷的袜子在某种程度上比无穷的鞋子多。

- 如果我能够获得永生，那么我可以一直磨磨蹭蹭。

- 你从甲地到乙地旅行，你需要经过两地之间的中点，然后经过剩余路程的中点，然后再经过剩余路程的中点，以此类推。剩余的路程永远有中点，那么你永远也到达不了你的目的地。是不是这样？

- 循环小数0.9999…等于1。

- 一个圆形是不是有无穷条边？

为什么数学好的人也会在微积分上卡壳？是的，这里面也有一个关于无穷的问题。

无穷能够通过不同的方式激发任何年龄段、任何知识水平的人的热情。这本书将会带领大家探索无穷并且超越无穷。如果你仔细思考并且使用正确的方式思考的话，你就会理解确实存在超越无穷的可能，就像我们总有更多的可问的问题和更多的值得探索的事物一样。无穷不是一个物理上的地点，所以我们要经历的并不是一个物理上的旅程。你可以坐在原地和我一起体验这个旅程，因为这个旅程是抽象的。这个旅程将会通往一个深入的、杂乱的、神秘的、无边无际的世界。

为什么

我们为什么要参与这个旅程？就像物理的旅程一样，抽象的旅程也有其存在的诸多意义。每个人都有自己的理由。也许你要在目的地做一件特别的事；也许终点处有非常好的风景；也许旅途中有美妙的景色；也许你喜欢行走或者攀爬所带来的物理体验，或者快速驾驶所带来的愉悦，又或者坐在火车上观看两边的田野向后飞掠所带来的静谧感（虽然我自己有关火车的经历往往牵涉延迟和恼怒的通勤者而不是静谧感，但是我们可以暂时把这些放在一边）；也许你喜欢探索未知的世界；也许你享受四处游荡，在陌生的城市迷路；也许你就是喜欢旅行，想要见识尽可能多的地方，因为“世界那么大，我想去看看”。

所有这些理由在抽象的世界中都有所对应。你要在目的地做一件特定的事——比如上班通勤，对应你脑中想要解决的一个特定的难题。这种抽象旅程的目的往往不是发现喜悦，而是完成任务。终点处的好风景对应我们通过研究获得的看待日常事物的新视角。旅途的美妙景色对应我们在研究过程中产生的神秘而美好的观点和看法。看着一个几乎不可能的观点逐渐被证实的喜悦，就像看到迷雾逐渐散去，大海开始在地平线上闪闪发光。我并不像很多人那样只是单纯地喜欢旅行，但是我有非常大的好奇心想要探索抽象的世界。我能够平静而顺从地接受这个世界上还有很多我没有去过的地方，但是当我遇到不能理解的观点时，我就会变得贪得无厌，想要探索一切。每当我瞥见自己不理解的事物时，我就会一门心思地扑上去。我喜欢让自己迷失在一个完全陌生的城市里，我也喜欢让自己迷失在一个陌生的观点里。虽然我竭尽所能地想要理解这些事物，但是我也乐于承认世界上的确存在一些人类无法理解的事物。事实上，我对此非常着迷。因为这意味着总有更多的事物存在。对我来说，这是一件美好的事情。如果有一天我们说“就这些了，没有更多了”，你不觉得这听起来有一些伤感吗？就像如果有一天，我说我已经尝遍了伦敦的每一个餐馆一样（当然这在事实上是不可能的）。总有一个你没有尝过的餐馆，同样的，总有一些我们

还不理解的事物。

从一个奇怪的角度看，这本书一点儿都不像是关于无穷的。它讲的是一个通往未知世界的惊喜之旅。儒勒·凡尔纳[注]的《地心游记》并不是讲地心的，它讲的是一个不可思议的奇妙旅程：抽象的思考是怎样进行的，以及这类思考能帮助我们做什么。当我们开始产生一个有趣的想法时，这本书可以帮助我们找到这个想法的本质。就本书而言，它可能并不会解释所有的事情——没错，数学家也无法解释关于无穷的每一件事——但是它可以帮助我们搞清楚，借助无穷这个概念，我们可以做什么，不可以做什么。

这本书的第一部分将帮助我们弄懂无穷是什么。如果你问一个小孩子无穷是什么，他可能会说"它比任何数字都更加大"。这话没错，但还是没有告诉我们无穷是什么。就像"姚明比你见过的其他人都要高"并没有告诉我们姚明是谁一样。

在这本书的第二部分，我们将利用关于无穷的观点来审视我们周围的世界，看看这个难以捉摸的怪物都到过哪些地方。它存在于我们用来玩乐的镜子里，存在于我们追跑打闹的路径中，存在于我们的每一段旅程里，存在于我们无穷变幻的世界的每一个情景中。理解无穷是进入微积分领域的基础，而微积分，毫无例外地存在于我们现代生活的方方面面。

所以，对微积分稍有理解将有助于我们享受当代生活的方方面面。但我并没有把无穷的这些现实应用当作这本书的主要写作方向。数学有一个很沉重的负担，就是实用性。诗歌、音乐或者足球就没有这种负担。如果你问我这些知识有什么用的话，我就会说，它帮助我们发电，打电话，建造桥梁、道路和飞机，向城市提供生活用水，开发药物，挽救生命。但这并不是说因为你思考了这些问题，所以你可以使用这些产品，而是因为其他人思考了这些问题，所以我们才能使用这些产品。因此，这些应用并不是我思考这个问题的原因，也不是我想要讲述这个故事的原因。

即便你在5岁之后就不再学习任何关于无穷的知识了，你还是能够生活得一样好。但是对于我来说，数学的价值并不在于"过日子"，而在于数学的思维和研究为我们的思索带来了光明。学习数学真正的目的是转过头来更好地理解我们的生活。就像在空中飞得更高能够让我们看到更远的地方一样。

让我们开始这个旅程吧。

---

1. 儒勒·凡尔纳（1828—1905），19世纪法国小说家、剧作家及诗人。——译者注

2　希尔伯特旅馆实验

数学可以被想象成很多事物：一门语言、一种工具、一个游戏。当你努力做家庭作业或者准备考试的时候，它可能不那么像是一个游戏。但是对于我来说，做研究最令人兴奋的一个阶段就是开始一个新的项目的时候。因为在这个阶段，你可以针对不同的观点进行有趣的尝试。这有点儿像在厨房里尝试不同的调味品。这比之后你努力记下发明出来的配方以免第二次做出来的味道不一样要有趣得多。当然，这比努力记下配方以免别人无法复制出来更加有趣。

所以作为开端，我会介绍一些关于无穷的有趣的观点，借此我们可以活跃一下我们的头脑，探索一下哪些关于无穷的观点可能是正确的，而这些观点的推论又是什么。数学的基础是通过逻辑理解事物。我们会发现，当我们对于“无穷”的理解不精确的时候，逻辑会把我们带到我们不曾设想的奇怪地点。数学家们往往会尝试从不同的视角来感受什么可能是正确的或者错误的。当乐高最初被设计出来之前，设计者一定尝试过很多不同的模型，而后才定稿完成最终的美妙设计。

一个数学家的“玩具”应该像乐高一样，足够强大，能够用来搭建事物，同时足够灵活，能够用来尝试不同的可能性。如果我们关于无穷的模型让一些基础概念出现了矛盾，那么我们就必须回过头去重新审视我们的模型。在最初的游戏之后，我们可能会多次回到我们的模型面前，因为我们会发现我们对无穷的思考可能引发了各种各样的问题。当我们最终得到了一个牢固的逻辑的时候，我们所完成的模型可能会和最初的想象完全不同。这同样会带来一些我们之前不曾预期的结果。比如，一个奇怪的事实就是，无穷可能有很多不同的“尺寸”。换言之，一些事物会比另外一些事物“更加无穷”。这正是所有旅程的美妙之处——发现一些不曾预期的事物。

在之前的章节中，我列出了一些关于无穷的基本观点，比如：

无穷会永远持续下去。

这是不是意味着无穷是一种时间、空间，或者长度呢？

无穷比最大的数字还要大。

无穷比我们能想到的任何巨大的事物都更大。

现在，无穷又有点儿像一种尺寸了。或者，它也许是一种更加抽象的事物——一个数字，我们可以用这个数字来测量时间、空间、长度、尺寸，甚

至任何我们想要测量的事物。接下来，我们会将无穷当作一种数字做进一步的研究。

无穷加一，它还是无穷。

换言之，

$$\infty + 1 = \infty$$

这看起来更像是关于无穷的基本原则。如果无穷是最大的事物的话，加上一并不会让它变得更大——真的是如此吗？如果我们在等式两边都减去无穷呢？如果我们用大家熟悉的消除法，在等式两边都消除无穷，那么等式就变成了：

$$1 = 0$$

这简直是一个灾难。一定有什么地方出错了。而下面的说法会导致更多的错误结果。

无穷加无穷，它还是无穷。

这看起来好像是说，

$$\infty + \infty = \infty$$

也就是说，

$$2\infty = \infty$$

现在，如果我们将两边都除以无穷的话，那么等式就变成了：

$$2 = 1$$

这成了另外一个灾难。现在，你几乎能猜到我们在思考最后一个观点的时候会发生什么。

无穷乘以无穷，它依旧是无穷。

如果我们把这句话写成一个等式的话，就是：

$$\infty \times \infty = \infty$$

如果我们在等式两边都除以无穷的话，就相当于在等式两边各去掉一个无

穷。等式就变成了：

$$\infty = 1$$

这可能是所有结果中错得最离谱的一个。无穷代表着最大的事物，肯定不会是1这么小。

到底是什么地方错了呢？问题就在于，我们像处理一个寻常的数字那样处理无穷，而我们并不知道是不是能这样处理它。我们在这本书中将会首先学到的事情之一就是无穷不是什么。我们会发现无穷肯定不是一个寻常的数字，继而渐渐了解无穷可以是什么。这个旅程花费了数学家几千年的时间，其中牵涉数学领域的很多重大的发展，集合论和微积分就是其中很好的例证。

上面的故事的关键在于，虽然无穷的概念很好建立，但是我们必须非常小心地处理它，否则就会发生相当奇怪的后果。而这些都仅仅是开胃小菜。我们接下来会看见各种各样的伴随无穷发生的奇怪的事物，比如事物的无穷集合、有无穷个房间的旅馆、无穷双袜子、无穷条路径、无穷多的点心。其中一些奇怪的发现就像“$1 = 0$”一样，不仅仅奇怪，而且让人不满意。所以我们需要自己构建数学模型来避免这些情况。但是，也有其他一些奇怪的事物并不违背逻辑，它们仅仅是违背常理。这些奇怪的事物并不会给我们的逻辑带来问题，却会挑战我们的想象力和思维方式，就好像科幻小说作家所塑造的那些拥有无穷生命、永生不老的人，或者拥有无穷的速度，能够瞬间移动的人一样。

拥有无穷多房间的旅馆

当我们开始教孩子们数字的时候，我们总会给他们一些实物帮助他们思考，或者我们会在他们吃一些可计数的食物时，教他们怎么计数，又或者，我们会教他们数自己吃了几勺子食物。

如果我们想要一勺一勺地数，一直数到无穷的话，那得花费非常多的时间。事实上，我们下面要介绍的几个例子确实有一点儿从一一直数到无穷的意味在里面。但是在做这些事情之前，我们还是先看一个已经是无穷的例子——一个拥有无穷多房间的旅馆。

想象一下，一个旅馆里面有无穷多个房间，房间的编号是1、2、3、4……直到无穷（见图2–1）。

1	2	3	4	5	6	7	8	9	10

图2–1

现在假设你是这个旅馆的经理。你面对的情况是每个房间都住了客人，而你正沉醉在你所赚到的钱里面。这个时候，另外一个客人走了进来，要求开一个房间。一方面，旅馆已经住满了。另一方面，如果你能让每一个客人都往后挪一个房间的话……

这个有无穷多个房间的旅馆被称作希尔伯特旅馆。德国著名数学家戴维·希尔伯特使用这个栩栩如生的例子来描绘你开始思考无穷时可能遇到的问题。一个正常的旅馆只会有有限的房间，住满了就是住满了。面对下一个客人，你根本就没办法安排，除非搭一个临时建筑。然而，在一个拥有无穷多个房间的旅馆中，你可以让1号房间的客人搬到2号房间，让2号房间的客人搬到3号房间，让3号房间的客人搬到4号房间，以此类推，我们总是能让n号房间的客人搬到n＋1号房间。因为我们有无穷多个房间，每一个n都有一个对应的n＋1，所以每一个客人都有一个对应的新房间。这么做的话，1号房间就空出来了，新的客人就可以入住了（见图2–2）。

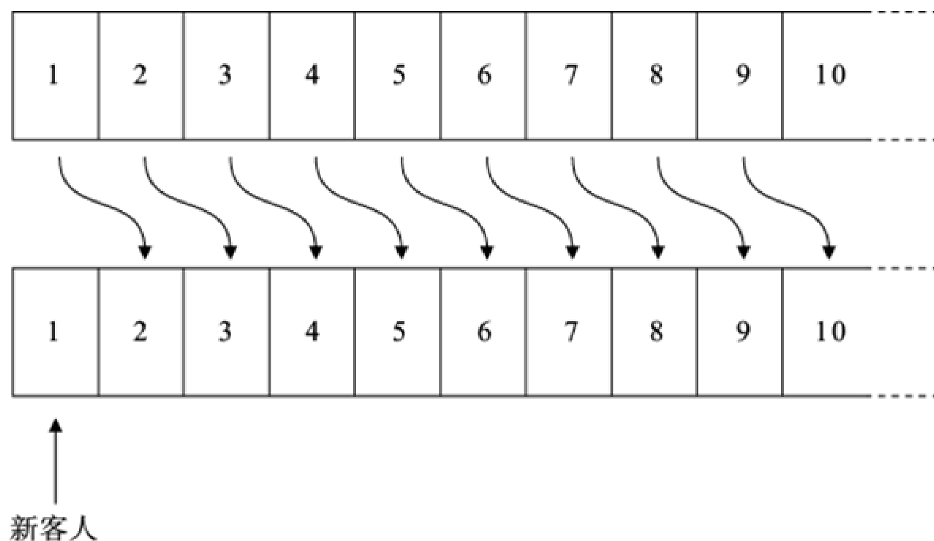


图2–2

这看起来是一个悖论，但是论证过程并没有漏洞。唯一的问题就是这个结

论与人们的直觉不相符。我们怎么能在已经完全住满的旅馆里再安排下一个客人呢？这和我们的直觉相悖的唯一原因就是我们太习惯于有限的旅馆了。当我们严肃地思考无穷，而不是模糊地想象无穷的时候，我们必须准备接受一些可能会显得有点儿奇怪的事物，甚至是看起来非常奇怪的事物。这也正是无穷的美妙之处。

我们想要做的是把“无穷”这个概念融入普通的数学中，而不改变其余的逻辑。就像科幻小说中永生不老的往往只有一个人，其他所有人都是有生老病死的普通人一样。一些奇怪的事情可能会发生，但是我们并不想因此而毁掉关于这个世界的一些基本事实。言下之意就是，我们并不希望因为将无穷和数学交织起来研究而发生“$1=0$”这样的事情。但是也许仍会有一些奇怪的新事物出现，就像这个拥有无穷多房间的旅馆一样。

希尔伯特旅馆并不会挑战现有的数学逻辑，它挑战的仅仅是我们关于旅馆的直觉。这个例子开拓了我们的眼界，让我们意识到，在无穷的情况下可能会发生的奇怪逸事。

如果来了更多的客人呢？

如果来了第二位客人呢？很简单，我们可以让每个客人都多往后挪一个房间。现在，原来住在1号房间的客人搬到了3号房间，原来住在2号房间的客人搬到了4号房间，原来住在n号房间的客人搬到了$n+2$号房间。这就是数学的世界，我们不需要考虑搬房间所带来的麻烦，我们只要开开心心地知道每个客人都有房间住就好了。

如果这两位客人同时到达，我们可以从一开始就让所有的客人都往后挪两个房间。当然，如果是三位客人同时到达的话，我们可以让每个人都往后挪三个房间。以此类推，只要是有限数量的客人同时到达，我们都可以用这种办法安排（见图2–3）。

如果有无穷多的客人同时到达怎么办？我们不能让每个客人都往后挪无穷个房间。虽然这个方案听起来好像有点儿道理，因为我们有无穷多个房间。但是让我们考虑一下某位特定客人的具体情况，比如1号房间的客人。这位客人要搬到哪个房间去呢？“$1+\infty$”号房间？这肯定不行，因为这就不是一个房间号。我们确实有无穷多个房间，但是每个房间还是有一个有限的房间号的。所以并不存在“$1+\infty$”号房间，让1号房间的客人搬到“$1+\infty$”号房间就等于这位客人还是没有地方可以去。如果我们不能告诉客人们他们应该搬到哪个房间去的话，那么我们就卡住了。

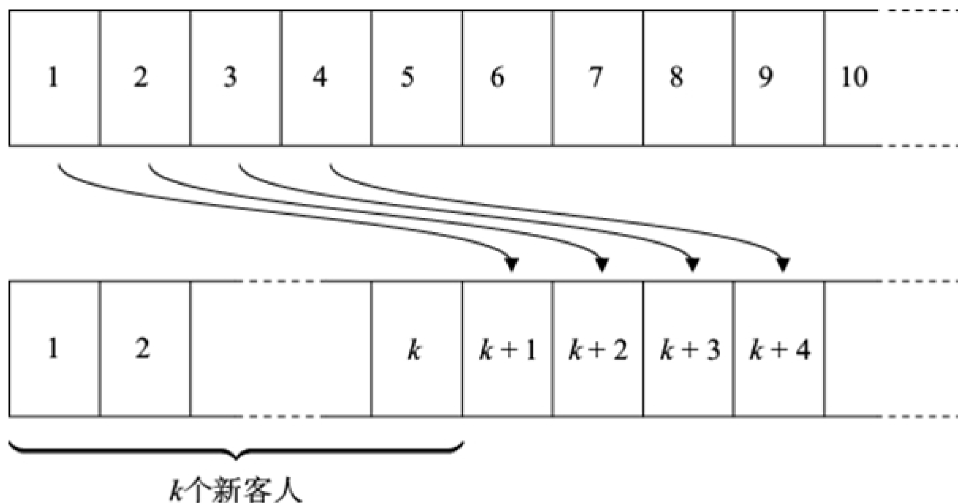


图2–3

所以我们不得不表现得更加聪明一点儿。（处理数学问题经常需要我们更加聪明，这也是数学看起来很难的一个原因。）我们可以让每个客人都去房间号是原来房间号两倍的房间。这样，1号房间的客人就去了2号房间，2号房间的客人就去了4号房间，n号房间的客人就去了2n号房间。（见图2–4）这样就空出来无穷多个房间。我们怎么会知道这样能行呢？我们知道本来已经入住的客人都已经搬到双倍房间号的房间了，所以他们现在全都住在偶数号的房间里。换言之，所有奇数号的房间都已经空出来了，而这样的房间有无穷多个。

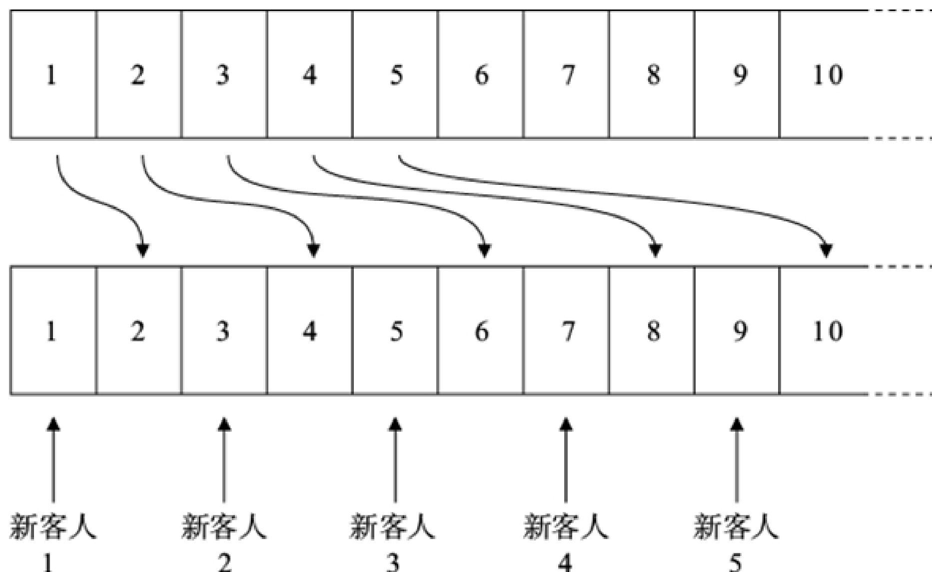


图2–4

事实上，我们可以写一个指导手册来告诉每位客人在不同的情况下他们接下来的房间号是什么。但是这个单子将会非常长，完成它花费的时间也会非常多。所以一个简便的办法就是我们可以写一个公式。使用公式的好处就是可以避免花费过多的精力写一个过长的清单。下面就是这个指导手册的简化版：

- 原来就已经在店里入住的客人：如果你住在n号房间，请搬到2n号房间。

- 新来的客人：如果你是第n号客人，请入住2n–1号房间。

现在，每个人都知道自己的房间号了。我们可以再检查一下，保证不会出现两个人被分配到同样的房间的情况，除非客人在计算的时候出现了问题。

你可能会注意到，这种情况只有在新来的客人已经排了队的情况下才能成为现实。否则，不守规矩的客人就会扭作一团，上演数学版的房间争夺大战。新来的客人必须按照编号顺序排队才能到达他们被分配的房间。因为现在情况变得复杂了，所以我们之后将会花点儿时间讨论一下队列的问题。

如果旅馆不止一层呢？

现在让我们假设我们的旅馆有两层，每一层都有无穷多个房间（见图2–5）。1楼有房间1、2、3、4……，2楼也有房间1、2、3、4……（更常见的编号方式是1楼有房间101、102、103、104……，2楼有房间201、202、203、204……，但是现在我们先不考虑这个问题）。

2楼	1	2	3	4	5	6	7	8	9	10
1楼	1	2	3	4	5	6	7	8	9	10

图2–5

如果这个旅馆起火了呢？现在，我们需要把所有的客人都转移到马路对面的只有一层的希尔伯特旅馆（这个旅馆刚好完全是空的）。这也不是什么难题。我们可以让原本住在1楼的客人把自己的房间号乘以二然后减去一，这样这些客人就分别去了1号房间、3号房间、5号房间、7号房间……，就像上一个例子中讲的新到的客人一样。接下来，我们会让原本住在2楼的客人都把自己原本的房间号乘以二，就像上一个例子里面原本就已经在店里入住了的客人一样。这些客人会住进2号房间、4号房间、6号房间、8号房间……（见图2–6）。

2楼 1 2 3 4 5

1楼 1 2 3 4 5

1 2 3 4 5 6 7 8 9 10

图2–6

从某种程度上讲，我们已经把“无穷×2”位客人装进了“无穷”个房间里。从数学上看，这和把新到达的无穷个客人安排进已经住满了的无穷多个房间的旅馆是一样的。

这个原则也可以应用到起了火的三层希尔伯特旅馆。唯一的不同就是，这次我们需要把“无穷×3”位客人安排进“无穷”个房间。所以我们需要把原本的房间号乘以三（见图2–7）。

- 原本住在1楼的客人需要把自己的房间号乘以三，然后减去二。那么他们就会住到1号房间、4号房间、7号房间、10号房间……

- 原本住在2楼的客人需要把自己的房间号乘以三，然后减去一。那么他们就会住到2号房间、5号房间、8号房间、11号房间……

- 原本住在3楼的客人需要把自己的房间号乘以三。那么他们就会住到3号房间、6号房间、9号房间、12号房间……

你可以想象一下，所有的客人按照他们原本的楼层排成了3列长队。然后你按照他们排队的次序安排房间，依次安排每个队伍的第一个人入住新房间。这里，你必须注意留好同一层客人入住房间的号码间隔，稍有不慎，你就没有足够的房间安排所有人了。

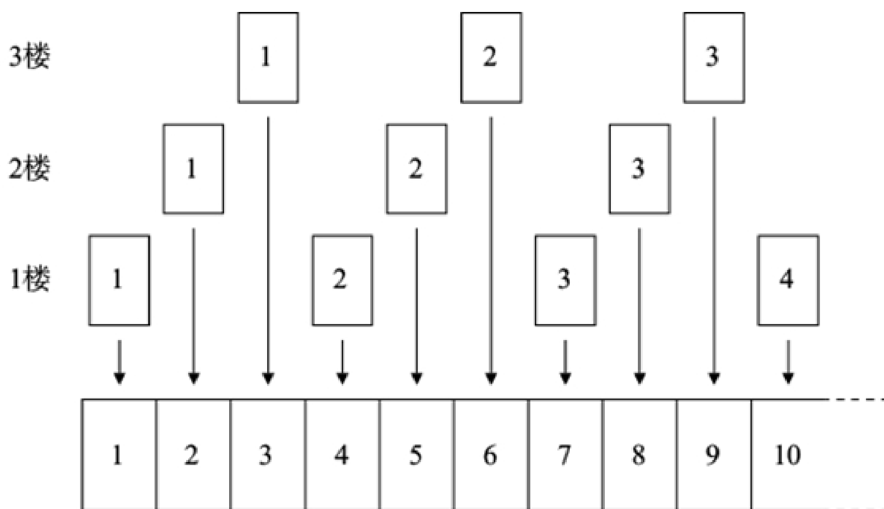


图2–7

像前面说的一样，我们可以为此写一个指导手册。正确的写作方式应该是下面这样的：

* 原本住在1楼的客人：如果你的房间号是n，那么请搬到3n–2号房间。

* 原本住在2楼的客人：如果你的房间号是n，那么请搬到3n–1号房间。

* 原本住在3楼的客人：如果你的房间号是n，那么请搬到3n号房间。

如果我们先安排所有原本住在1楼的客人的话，我们就会说：

* 原本住在1楼的客人：如果你的房间号是n，那么请搬到n号房间。

但是这样一来，我们是不是就没有房间安排原本住在2楼和3楼的客人了？

是的，已经没有了。因为每一个房间n都已经被原本住在1楼n号房间的客人占据了。这就是为什么我们要么得按照楼层顺序轮换安排客人，要么得在安排1楼的客人的时候给2楼和3楼的客人预留下房间，而不能先把1楼的客人按照原本的房间号安排进旅馆。

我希望你能够按照这个逻辑处理更多楼层的情况（见图2–8）。

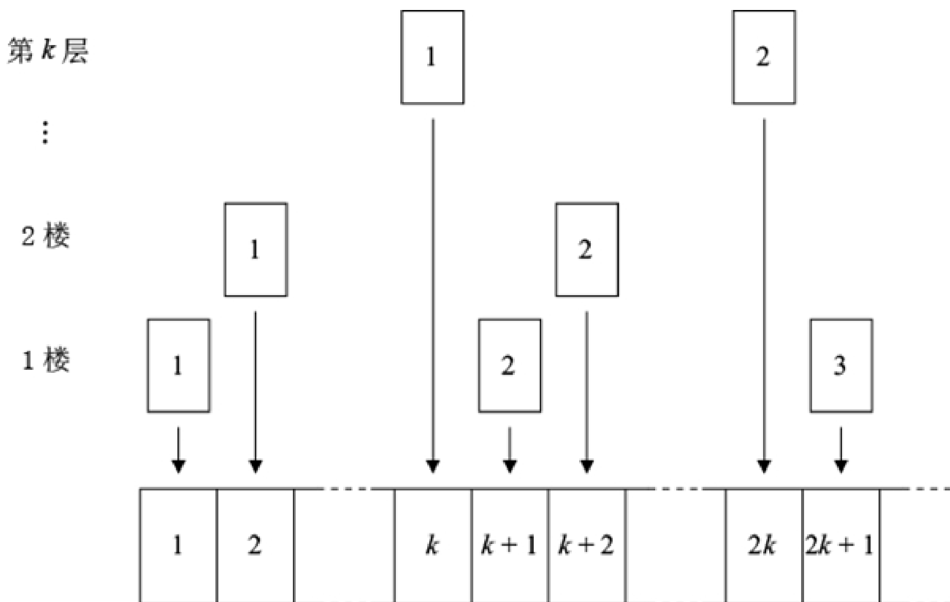


图2–8

但是如果有无穷层楼呢？现在，我们把希尔伯特旅馆想象成一座摩天大楼，楼层有第1层、第2层、第3层、第4层……，每个楼层都有1号房间、2号房间、3号房间、4号房间……。我们可以把这个建筑想象成“无穷乘以无穷”（见图2–9）。

⋮											
5楼	1	2	3	4	5	6	7	8	9	10	…
4楼	1	2	3	4	5	6	7	8	9	10	…
3楼	1	2	3	4	5	6	7	8	9	10	…
2楼	1	2	3	4	5	6	7	8	9	10	…
1楼	1	2	3	4	5	6	7	8	9	10	…

图2–9

如果这一回是这座摩天大楼起火，我们是不是就无计可施了呢？我们能不能把这栋摩天大楼里的客人转移到只有一层的希尔伯特旅馆呢？也许在现在的情况下，只有一层的希尔伯特旅馆看起来已经成了一个相当普通的概念。当我们一次又一次地锻炼我们的头脑的时候，就会出现这样的情况：原本非常令人诧异的事情变成了普普通通的事情。这标志着我们已经更加聪明了。

回到正题，你可能觉得这次的情况有点儿无望了，因为我们不能让每一个人都“把自己的房间号乘以无穷”，然后再减去点儿什么东西。我们也不能依次安排每一列楼层队列的第一个人了，因为如果我们这样做，就会发生下面的情况：

* 原本住在1楼1号房间的客人搬到1号房间

* 原本住在2楼1号房间的客人搬到2号房间

* 原本住在3楼1号房间的客人搬到3号房间

* □

* 原本住在n楼1号房间的客人搬到n号房间

* □

把每个楼层1号房间的人安排完之后，一层的希尔伯特旅馆就客满了。因为每一个n号房间都已经被原本住在n楼1号房间的客人占据了。

然而，我们并非完全没有办法。我们需要变得更聪明一点儿。问题的关键还是让大家排起队来。但是，这次我们让所有的客人按照对角线的方式排队（见图2–10）。

第5队列										⋮
第4队列	1	2	3	4	5	6	7	8	9	10	5楼
第3队列	1	2	3	4	5	6	7	8	9	10	4楼
第2队列	1	2	3	4	5	6	7	8	9	10	3楼
第1队列	1	2	3	4	5	6	7	8	9	10	2楼
	1	2	3	4	5	6	7	8	9	10	1楼

图2–10

如果我们从左下角开始，按照对角线的方式排队，我们还是能够把每一个客人都安排到新的房间。这次的安排方式无法像前几次那样用一个简单的公式总结，我们用一个图形来表示（见图2–11）。

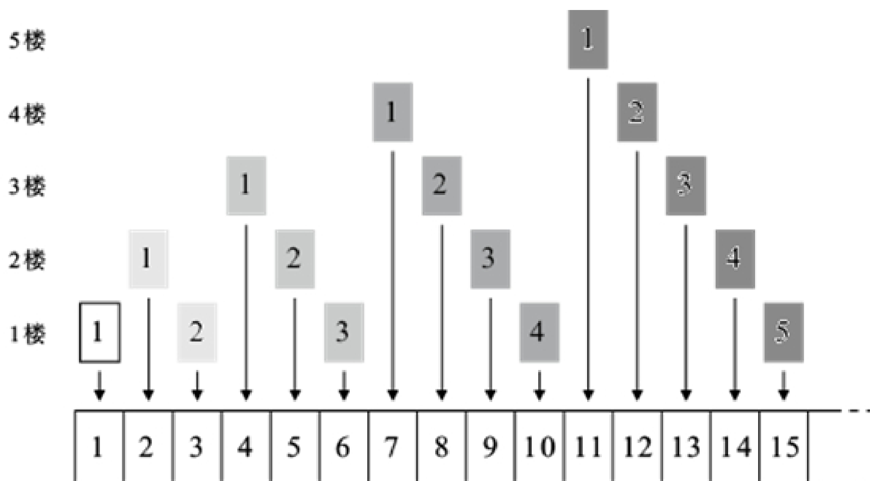


图2–11

一个指导每位客人应该去哪个房间的指导手册的文字描述需要像下面这样写：原本住在k楼n房间的客人搬到……房间。你可能可以根据上面的图总结出一个公式来，但是我觉得在这次的情况下，用图来表示会更加清晰一些。

顺带一提，我们关于希尔伯特旅馆的讨论包含了下面这个奇怪的事实：偶数的个数和所有数字的个数一样多。因为当你让每一位客人都把自己的房间号乘以2的时候，你就用偶数房间号的房间安排下了所有的客人。而我们将一整层的客人安排到奇数房间号的房间的事实也说明奇数的个数和所有数字的个数一样多。按照这个逻辑，如果我们有无穷多的钱的话，我们能表现得极为博爱——我们可以把无穷多的钱捐给慈善机构而自己仍然剩下无穷多的钱。我们需要做的就是把银行里的每一块钱中的偶数号捐给慈善机构，自己留下奇数号。但是这显然不太现实，因为银行里的钱并没有编号，有的只是一个总数。但是我们可以转一块钱到慈善账户，再转一块钱到自己的个人账户，然后再转一块钱到慈善账户，再转一块钱到自己的个人账户。这做起来有点儿慢。所以你也可以一次转一亿元到慈善账户，再转一亿元到自己的个人账户，以此类推。但是，你需要一直不停地这么做下去。

受到成功完成这个几乎不可能完成的任务的鼓舞，你可能会觉得你现在可

以把任何旅馆的全部客人转移到仅有一层无穷多房间的希尔伯特旅馆里了。然而，事实并非如此。如果你有另外一个更加疯狂的旅馆，旅馆的房间编号包括所有的有理数和所有的无理数（“你好，我在π房间”），那么我们可能就真的被打败了。所以，事情的关键在于“可数性”。我们接下来就会开始接触这个概念。在第6章，我们会看到一个令人脑洞大开的事实，即，有一些无穷比另外一些无穷要大。

无穷令人着迷的一点就是你总会在无意间撞见这个概念，而且总会无意间撞见围绕着这个概念发生的神奇的事情，但是要搞清楚这些事情背后的原因则非常困难。我们现在知道，一个有无穷多的房间的旅馆和“常规”的旅馆非常不一样。我们还知道，我们不能像处理“常规”的等式那样处理涉及无穷的等式。看起来，无穷好像不是一个“常规”的数字，那么它到底是什么？数字看起来是数学的基石，但是数字到底是什么？有很多的道理我们认为是理所当然的，但是我们从来没有思考过它们的本质是什么，数字就是其中之一。如果我们想要宣称无穷不是一个数字的话，我们最好先搞清楚数字是什么。你可能会很诧异地发现数学家竟然花了如此长的时间来搞懂数字的本质。人类虽然并不清楚数字是什么，却还是毫无障碍地使用了它上千年的时间，你可能也是如此。有鉴于此，你可能会觉得弄清楚数字的本质是一件毫无意义的事情。那么，难道说数学家在做这件事情的时候是非理性的吗？

事情是这样的。常规的整数是不难理解的。即便你将常规的数字扩展到负数和分数的领域，也不是什么大问题。问题出现在分数与分数之间这个领域，也就是无理数，此时事情就开始变得难以捉摸起来。不明白整数是什么不是什么大问题，但是不明白无理数是什么就成了问题。在移除这个障碍的过程中，微积分出现了，而微积分极大提升了过去两个世纪里科学、医药学和工程学领域的精确度和人们对其的理解。而为了更好地理解这些无理数，我们需要更好地理解所有的数字，包括最基本的数字。我们需要将地基打好，如此才能在其上构建坚固的建筑。如果地基不稳，那么除了回头去重新打好基础之外，我们别无他法。

3 无穷不是自然数

当孩子们第一次学会爬一小步的时候，他们对此完全着了迷。他们会一遍一遍地做这个动作，爬得越来越高，越来越高。他们会一直爬到楼上去，除非担心的大人把他们拦住。类似地，在学会了数数之后，他们就会不停地一个数字一个数字地加上去。但是通过这种方式，他们是不是能够到达无穷呢？

要回答无穷是不是一个数的问题，这是第一步。这就像那个荒唐的法律案件，法官要搞清Jaffa Cake（佳发）到底是蛋糕还是饼干（因为在当时的英国法律系统里，蛋糕和饼干的税率不同）。如果你不先坐下来给“蛋糕”和“饼干”各下一个定义的话，你是没办法回答这个问题的。Jaffa Cake的小尺寸和扁平形状使其偏向于饼干，但是它的纹理及松软的质地则使其偏向于蛋糕。Jaffa Cake会像蛋糕一样在变质了以后就变硬，而不是像饼干一样在变质以后就变软。这个案件最终的判决注重考量Jaffa Cake的纹理和变质特性，法官最终将Jaffa Cake划定为蛋糕，适用比较低的税率。

那么无穷呢？让我们想象这样一个法律案件，案件的关键在于无穷到底是不是数字。那么我们要怎么定义数字的特性呢？这就是我们在这一章里要开始思考的。事实上，存在很多种不同的数字。从最简单的1、2、3，到负数、分数、无理数和一些更加奇怪和美妙的数字。我们会一步一步地揭示，无穷不属于这些类型中的任何一种。

你可能会很奇怪，为什么我们不能直接把无穷定义为一个数字，然后按照处理数字的方式去处理它呢？为了弄明白这个道理，我们需要搞懂数学的逻辑。这就好像你在字典里查一个词的时候，这个词的定义包含另外一个你不认识的词；而当你查这个新的词的时候，另外一个新的词又出现了。为了弄明白无穷，我们需要弄明白数字，而为了弄明白数字，我们需要弄明白数学。等一等，为了弄明白数学，我们还要搞清楚逻辑。

数学使用逻辑研究事物，所以数学所研究的只能是符合逻辑的事物。当我们定义数学对象的时候，我们可以采取各种各样的方式。我们可以通过罗列来说明它们是什么，我们也可以通过定性来说明它们是什么。我们可以通过罗列世界上所有的鸟类，写一个长长的单子来说明什么是鸟，也可以用“鸟是有羽毛、翅膀和喙的动物”来说明什么是鸟。第二种方法不但更快，而且为你目前还不认识的鸟类留下了充足的补充空间。此外，当你遇到一种新的生物的时候，你还能利用这个定义来判断它是不是鸟类。

下面是一个数学的例子。我们可以说，“正多面体包括正四面体、正六面

体、正八面体、正十二面体和正二十面体”。我们也可以说，“正多面体是这样一种立体形状，其所有的面、边和角都是相同的”。或者更加精确地说，“常规的正多面体，其各个面都是全等的正多边形，并且各个多面角都是全等的多面角”。在第一种情况下，我们罗列出了所有的正多面体。由于其总量不大，因此这么做是可行的。但是在第二种情况下，我们描述了所有这些正多面体所具备的特性。根据这个定义，人们可以在不用理解什么是二十面体的情况下，判断一个物体是不是正多面体。

对于一些数学对象，罗列是很困难的，因为就像鸟一样，它们所包含的个体数量太大。比如，我们不可能罗列所有的数字，因为其个数是无穷的。我们也不可能罗列所有的质数，这里的原因有很多。首先，我们不知道所有的质数都有哪些。其次，即便我们知道，我们还是无法一一罗列，因为其数量是无穷的。在质数这个例子里，我们可以将其定性描述为“只能被1和自身整除的数，而1不是质数”。然后，我们面临的挑战就是找出所有满足这个条件的数来做列表。

为了证明一些事物不是某个特定的数学对象，我们通常也有两种方法。我们可以盯着那张关于该数学对象的清单证明该事物不在清单上，或者我们可以参照该数学对象的特性证明该事物不具备这些特性。比如说，我们可以轻易地看出球体不是正多面体，因为它不在正多面体的清单上。但是我们不能通过比对清单证明一个数字是不是质数，因为这个清单并不存在。因此，我们只能去检查该数字是不是符合质数的特性。换言之，我们需要验证该数字是不是只能被1和其自身整除且不是1。比如说，6还能被2整除，所以它不是质数。如果我们在森林里发现一个新的生物，并且想要判断它是不是鸟类，我们需要检查它是不是符合鸟类的特性。在某个时刻，生物学家不得不去定义鸟类的特性，而数学家也不得不去定义数字的特性。

事实上，在一定的数字范围内，质数清单是存在的。只要你的数字不是特别大，你就可以通过比对清单来确定你的数字是不是质数。然而，计算机需要一种方法能够快速确定一个数字是不是质数，而不是耗费大量的空间存储质数清单。截至我写这本书的时候，人类确定的最大的质数大概有2200万位数字。这个数字是如此之大，以至于仅存储这个数字一件事就几乎是不可能的了，何况还需要存储小于这个数字的所有质数。所以，事实上，并没有人会保存这样一个清单来记录所有的质数。

就像鸟类一样，世界上有各种各样的数字。一些比另一些更普遍，而一些比另一些更加广为人知。我们将会发现，最常见的数字并不是我们最常思考和讨论的。我们先来研究一下最显而易见的数字，也是我们最常思考的数字：我们用来计数的数字。

自然数

最基本的数字就是小孩子们最初用来数数的数字：1、2、3、4……。在数学领域，这些数字被称为自然数，因为它们是最自然的。我刚刚描述这些数字的方式就表明了这一点——简单地数1、2、3、4……。但是，如果我们从来没有学过怎么数数，我们怎么能知道应该怎么往下数呢？我们怎么知道4的下一个数字是5，而不是$4\frac{1}{2}$？我们怎么知道$4\frac{1}{2}$不是这个单子里面的数？

这其实是一个意义深远的问题。我们知道4的下一个数字是5，是因为我们宣布了下一个数字是5。但5是什么呢？我们怎样才能不依赖为每一个数字命名的方式来定义自然数呢？

数学家的定义方式基本上就是说，如果我们继续数下去的话，下一个数到的数字就是它。但是我们怎么才能说得更加精确呢？我们需要说得更精确，因为我们需要确保它遵从逻辑。“如果我们继续数下去的话，下一个数到的数字就是它”，这种说法并不符合逻辑，因为我们根本并不知道“继续数下去”是什么意思。我们要把这句话说得符合数学逻辑，不能模棱两可。

所以，我们不说“继续数”而是说“加上1”。自然数就是我们在前面一个数字上加1所获得的数字。但是我们要怎么开始呢？我们必须从一个地方开始，否则我们什么数也数不到。于是，我们决定从1开始。

自然数就是我们从1开始不停地加1所获得的数字。事实上，对于0是不是自然数这个问题，人们有不同的看法。但是这并不重要，因为我们也可以从0开始不停地加1。

有些人对于0是不是自然数这件事看得很重，而我觉得这不过是一个文字和名称的游戏。从数学的角度讲，你可以从任何地方开始，然后不停地加1。一些数学家认为从0开始很好，因为0是一个非常有用的数字。但是另外一些人觉得，0不能被称为一个“自然数”。我认为，这是一个语言学问题，而不是一个数学问题。但是这一说法也是有争议的，有人认为定义学是数学的一个部分。我个人认为关于0是不是自然数的争论只会把人们带离数学本身。但还是有一些人觉得，这件事非常重要，他们会像我发送不友好的电子邮件要求我修正我的说法。有几次，我在演讲过后还遇到了有人与我争论这个问题的情况。

如果我们认定自然数就是“从1开始不停地加1所获得的数字”，那么自然数

就可以被这样描述：

$$1$$

$$1+1$$

$$1+1+1$$

$$1+1+1+1$$

$$1+1+1+1+1$$



因为一直这么做的话会非常枯燥无味，所以我们给了这些数字新的名字。这在某种程度上是一个简称，让我们可以不用说“1加1加1加1”，而可以直接说“4”，这样会便捷很多。数学语言给了我们简便的办法来称呼冗长的事物，但它看起来也可能像是令人困惑的行话。你可能不会觉得“1、2、3、4”像是令人困惑的行话，那是因为在你使用了足够长的时间之后，行话变得让你熟悉了。如果你的工作和我的有一点儿像的话，那么你在开始接触这个工作时，很可能会因为听不懂人们使用的术语而感到头昏脑涨。但是几个月之后，你也能熟练地使用那些术语了。

当你开始一份新工作的时候，你可能会收到一张清单，上面列着你需要熟悉的术语。我们不可能为每一个可能的数字创造一个简短的名字，但是我们可以设定一个原则用来给新的数字命名。你在学习外语的时候可能也接触过这个办法。你通常要通过死记硬背记住1到10这10个数字单词的拼写，而从10到20之间的某个位置开始，你就会发现一定的规律了。在英语里，除了个别地方的拼写有一定的特殊性之外，从15（fifteen）之后，16（sixteen）、17（seventeen）、18［eigh-teen］到19（nineteen）都遵循一定的读写规律。甚至如果你把拼写差异放宽的话，13（thirteen）、14（fourteen）、15（fifteen）也或多或少遵循着同样的规律。在西班牙语里，从16开始就有一定的规律了，而在法语里，你需要等到17，在德语里你只要等到13就可以了。在中文里，从11就开始有规律了，因为“十一”就是一个“十”加上一个“一”，“十二”就是一个“十”加上一个“二”，以下同理。

一旦你学习到了20，通常你就只需要再学习20、30、40、50、60、70、80、90和100的写法就可以了。这些数字的写法也都遵循着一定的规律，在不同的语言里，这些规律可能略有不同。在英语里，这些数字的写法大致遵循一定的规则，但是会有一些特殊的变化。在德语里，20之后的数字

单词的拼写规律非常整齐。在法语里，有一些奇怪的现象可能会让外国人觉得疑惑，因为从70开始，你就要数“六十十”“六十十一”“六十十二”了，90以上也是如此。这里，瑞士法语是一个例外，因为瑞士法语里有一个专门的词用来表示70，就是“septante”。

中文再一次在便捷性上领先，因为“二十”“三十”“四十”都是非常简单直接的表达。我从印地语讲师杰森·格鲁内鲍姆那里学习到，在印地语里，从1到100都分别对应一个单独的词。鉴于此，他会给能够背下来所有这100个词的学生额外的奖励分数。

在100之后，基本上你只需要记住每一个数量级所对应的词就可以了。比如英语里的千（thousand）、百万（million）、十亿（billion）、万亿（trillion），或者中文里的千、万。英语里没有中文中“万”这个数量级，其对应的说法是“十千”。

除此之外，我们就没有新的词了。而我们用到更大的数字的机会也非常少。人们很少会用到一千万亿兆以上的数字。我几乎从来没有遇到过需要说这样的数字的机会。（除了在想要表述数量大得夸张的时候。比如，我会说美国的大学花费要“上千万亿”甚至“上兆亿”。）

但事实上，更大的数字的确存在，只不过我们并没有专门为其命名。就像一些动物无论人类有没有为之取名它们都存在一样。一个更好的例子是海王星，人们在发现和命名这个星球之前，就已经通过数学的方法推导出这个星球的存在了。我们知道，超过一万亿的数字是存在的，因为我们知道我们可以不停地加一。事实上，我们可以做一个简单的实验来证明超过我们的命名系统的数字是存在的。

假设A是我们能够命名的最大的数，我们总是能够找到A＋1这一更大的数。

这让我想起电影《山村犹有读书声》里面的一个感人场景。杰出的教师乔治·洛佩斯在法国乡村一所学校里教一个班的小男孩，男孩的年纪从4岁到10岁不等。影片中，一个小男孩做了一些淘气的事情，搞得自己满手都是墨水。洛佩斯把他带去洗手，没有责备他，反而开始问他一些数学问题。他不停地问这个小男孩最大的数字是多少，而小男孩总是斩钉截铁地给出一个答案——他很确定自己知道最大的数字是多少。开始的时候，他认为100是最大的数字，之后教师温柔地问他“那101呢？”这个过程不断重复，一直到你看见小男孩的眼睛因惊奇而瞪得越来越大，因为他意识到这个对话可能会一直持续下去。

不幸的是，这部电影的巨大成功引发了来自电影主角的原型乔治·洛佩斯先

生的法律控诉，因为看起来，电影的制片人利用洛佩斯先生杰出的教学方法赚了很多钱。对于这位老师应该从中分得多少报酬这件事，大众的观点产生了分歧，而法院的判决结果是：洛佩斯无权分得电影收益。一些人批评洛佩斯过于贪财，破坏了电影中那位无私的教师给他们留下的美好印象。我对此感到很伤心——竟然有人认为一个投身于教学事业并改变了很多人命运的教师并不值得获得金钱的回报。

无穷不是自然数

到现在为止，我们知道了我们可以从1开始不停地加1从而获得所有的自然数。但是，我们怎么知道这么加下去我们不能到达无穷呢？为了回答这个问题，我们需要把自然数的定义描述得更加精确一点儿。

自然数是所有这些数字的集合，这些数字有两种可能性：

- 1

- n＋1，其中n自身也是一个自然数

所以，2是一个自然数，因为2＝1＋1。同样，3也是一个自然数，因为3＝2＋1，而2是一个自然数。为了证明10是一个自然数，我们需要重复这一步骤9次。虽然听起来有一点儿乏味，但这是我们能够做到的。

那么，我们能用这个办法证明∞是一个自然数吗？首先，∞肯定不是1，所以排除了第一种可能性。其次，是否存在一个自然数n能够满足∞＝n＋1这个条件呢？问题来了，要满足这个条件，n需要是∞–1，而∞–1仍然是∞。所以这个命题就等同于：如果∞是一个自然数，那么∞就是一个自然数。这是一个循环论证，对我们一点儿帮助也没有。

经过上述努力，我们还是没能证明无穷是一个自然数。我们只证明了无穷不是一个显而易见的自然数。但是，让我们停下来思考一下我们试图做的这件不太可能的事情：我们并不知道无穷是什么，而如果我们不知道一个事物是什么，我们就很难证明它不是什么。

等一等，真的是这样吗？

我们不知道怎么定义无穷，但是我们知道，如果我们可以定义无穷，那么其定义一定会包括这几个条件：

* 加上1不会让它变得更大。

* 加上无穷也不会让它变得更大。

* 乘以一个自然数也不会让它变得更大。

我们接下来将要证明自然数不可能满足上述无穷定义中的这些条件。

我一直认为，数学对象的特点是，只要你能想到，它就是存在的，唯一的前提是它不会造成什么悖论。在这里，我们“想象”出了一个叫作无穷的数学对象。在我们的想象里，它满足我们描述的这些规则。为了证明其存在，我们需要证明它不会造成任何悖论。不幸的是，它确实会造成悖论！我们将要通过这种反证的方式证明无穷不是一个自然数。也就是说，我们可以先把它想象成一个自然数，然后证明这么做会带来显而易见的错误。

现在，我们得先了解一些关于自然数的原则。比如说，

* 无论你用什么顺序将自然数相加，你得到的结果总是一样的，比如$3+2=2+3$。

* 你可以让自然数们相减（注意不要产生负数，因为我们还没有介绍到负数）。

* 如果你在等式两边同时加减一个数的话，等式仍然成立。

现在，我们将这些原则应用在无穷上。首先，我们看一下无穷的第一个特征。起始等式如下：

$$\infty + 1 = \infty$$

现在我们在等式两边同时减去∞，就会得到：

$$1 = 0$$

很显然，出错了。我们刚刚尝试了将无穷代入一个自然数的原则——在等式两边同时减去一个数后，等式仍然成立，而结果是错的。这里符合逻辑的总结就是，同时在等式两边减去无穷会导致等式不成立，既然在等式两边同时减去一个自然数后，等式仍然成立，那么我们可以得到结论：无穷

不是一个自然数。

这是不是有点儿像循环论证？

想象一下，你和小孩一起玩游戏，你充满爱意地说，“你是我最喜欢的小兔子！”结果却遭到这个孩子的反驳，“我不是一只兔子！”现实中，很多大人都喜欢和小孩子进行这种毫无意义的、傻乎乎的，却很可爱的对话。在上面这种情况下，大人就可能会说，“你当然是一只小兔子！”这时候，孩子则会抗议说，“但是我没有毛茸茸的小尾巴！”

他们所做的事情就是反证的一个简单的例子。

假设我是一只小兔子。

那么我就会有一个毛茸茸的小尾巴。

但是我没有一个毛茸茸的小尾巴。

所以我不是一只小兔子。

同样，我们也以此证明了无穷不是一个自然数。

假设无穷是一个自然数。

那么我们就可以从等式两边同时减去无穷，并使等式仍然成立。

但是我们不能从等式两边同时减去无穷，并使等式仍然成立。

所以无穷不是一个自然数。

但是，这并不代表无穷不是一个数字，只是说明了无穷不是一个自然数。我最近被叫去为我4岁的外甥和他的好朋友之间的一次争吵做裁判。他的好朋友说，“无穷不是一个数字，因为我爸爸是这么告诉我的。我爸爸是一个科学家，他什么都知道。”而我的外甥很聪明地集中火力攻击其中最容易推翻的论据，那就是“他好朋友的爸爸什么都知道”这一点。然而我则一直试图说服我的外甥，数学家们知道一些科学家们并不知道的事情……

4　有理数和无理数

“我们到了没有？”这是小孩子在长途旅行里最常问的一个问题。他们对于时间的感受和成年人不同，10分钟对于他们来说可能已经是很长的一段时间了。所以，如果这段旅程要花费几个小时的话，作为家长的大人们就要做好心理准备了。

有一次，我沿着密歇根湖湖岸散步，在到达一个沙丘后，我决定从此处沿岸边游回出发地。起起伏伏的水面遮挡了我的视线，我总觉得我肯定已经离出发地不远了。但是每次靠岸时，我总会发现目的地还很遥远，湖岸线仿佛无穷延伸下去一般。最终，我不得不通过在脑子里唱歌来记录我到底游了多久。

有的时候，数学看起来就像是一个永远到不了任何地方的旅程。因为每当你刚刚弄清楚一个新的事物的时候，就会发现它又带来了很多你还不明白的事物。而且，我们对自己从哪里来、到过什么地方的记忆也无法一直保持下去，因为一旦我们弄明白了一个问题之后，就想不起来这个问题曾经有多么难。我经常会觉得我在数学上什么成果都没有取得，因为我已经弄懂的问题都显得如此简单，而我还没有弄懂的问题都显得如此困难——若非如此，我早就应该弄懂它们了。

我们在前一章证明了自然数无法将我们带到无穷。这一章，我们会继续我们的数字之旅。孩子们在掌握了如何数数之后，往往会立刻开始学习如何“反着数”，换言之，如何做减法。随着孩子们对数字了解得越来越多，他们也渐渐接触到了越来越多的数字类型。某种程度上，这个过程有点儿像数学家发现数字的历史。换言之，几千年的数学变迁被浓缩成了几年的学校教育。这就是教育的力量。

但是，我们需要仔细地交叉比对这些数字类型，才能判断出无穷是否属于其中的某一类。像很多事情一样，如果你始终不问及更加困难的那些问题，你将永远无法获得更深层次的理解。“什么是无穷？”就是这类更加困难的问题中的一个。花这么多的时间来思考一个定义可能会显得无用、枯燥，甚至迂腐，但是我更喜欢把这看作一个揭示真理的过程，一个不断验证自己的想法是否正确的过程。如果我在一家餐馆吃了一顿美味大餐，我立刻就想知道这些菜是怎么做的。如果我参与了一趟美妙的旅程，我就希望能够在地图上看到我走过的路。研究这些数字到底是什么就是一个类似的过程。我们将会看到每一个新的数字类型是怎样在我们已经知道的类型的基础上构建出来的，以及这一过程展现了数学拥有怎样的精妙性及表现力。

用旧数字构建新数字

数学家们喜欢在旧的事物的基础上用尽可能少的力气构建新的事物。这听起来是一种很懒的办法，但我更倾向于认为这是一种节省脑力的做法。我们的大脑容量是有限的，因而我们的脑力也是有限的，我们要把脑力集中用于我们需要它的地方。

有一个关于数学家的笑话是这样的：

第一次，你给数学家一个炖锅和一个鸡蛋，让他们煮鸡蛋。他们会把炖锅里装满水，放进鸡蛋开始煮。第二次，你给数学家一个装满水的炖锅和一个鸡蛋，同样让他们煮鸡蛋。他们就会把水倒掉，然后说，“我把问题的难度降到了和上次一样的水平”。

利用以往的经验来构建新的事物除了节省脑力以外还有很多其他的好处，比如帮助我们看到不同概念之间的关系，理解不同的事物是怎么结合在一起的。想象一下火焰雪山（Baked Alaska）注的食谱，这个食谱并没有告诉你最终的作品是一个蛋糕、冰激凌和蛋白酥皮的混合体。那么，如果你之前并不知道火焰雪山是什么的话，整个制作过程会让你感到非常困惑。

从自然数开始，我们会构建出越来越复杂的数字类型。这些数字会显得越来越不自然，不自然到获得“无理数”和“虚数”这样的名字。数学家们喜欢用日常词汇来命名新的概念，从而让我们更好地理解这些概念的特性。如果我们赋予抽象的概念一些我们能够联想到其特性的实物名称的话，我们就能更加容易地理解这些概念。我觉得“虚数”（imaginary number）这个名字就显得很讨人喜欢，而“素理想”（prime ideal）这个概念听起来就像一块儿多汁的牛排。当然，也许只有我自己这样认为。需要说明的是，“素理想”这个概念和本章的主题并没有关系，只是我个人觉得这个名词听起来非常美味。

我们在自然数的基础上构建出的下一个概念是整数。整数包括所有的自然数及其所对应的负数，还有零。当然，是否包含零取决于我们是否把零算进自然数的范围内。整数的概念来自我们对于减法的喜爱（如果这种喜爱真的存在的话）。你可能还记得小时候你在还没有接触到负数的时候所面临的纠结。你能够把任何的数加在一起（至少在理论上是可行的），但是你只能用大的数减去小的数。同样，数学家也曾经为此而困扰。为了能够用任何数减去任何数，我们需要自然数的负数形式，由此就产生了整数这个概念。整数比自然数“好”上一点点，因为它能够帮助我们做更多的事。

数学的发展往往起因于数学家在现实世界因无法计算某些事物而感到挫败。所以，他们发明了一个新的概念，构建了一个新的世界，在这个新的世界里他们就能做他们想做的事情了。我喜欢把我们这些数学家想象成顽固的规则破坏者。当我们在面对一条规定我们不能做什么事情的规则时，我们就会思考：我们是不是可以构建一个新的世界，而在这个新的世界里，我们就可以做这件事情了。这和大众对于数学这个学科的理解非常不同。在常规的理解里，学习数学意味着必须遵循一大堆各种各样的规则而非打破规则。

如果说我们的自然数系统可以不包含零，那么我们在整数的世界里最需要的事物就是零。零是一个特别的数字，它的特性是：在任何一个数字上加上零之后，什么都不会发生。

$$0+1=1$$

$$0+2=2$$

$$0+3=3$$

……

$0+n=n$，无论n是什么

……

下一步，我们希望加法能够“逆转”或者“撤销”。这就像是你在买东西的时候希望商店允许退货一样。如果我买了什么东西以后不能再因改变主意要求退货的话，我就会有点儿紧张。这不代表我总会改变主意并要求退货，只是知道我可以这么做就能让我感到放松。

一旦你加上了一个数之后，你怎样才能改变主意并且“退回”加上的那个数字呢？你可以减去这个数字。另一个思考角度是，你可以加上一个负数。就像是你在店里退了一件货品然后拿到一张写着负值的收据一样。你会收到一条信用卡刷了一个负值的信息，其实际含义就是银行柜员从你的账单上减去了这个值。

对于我们来说，多花点儿时间理解减法是很必要的。因为我们在上一章考察无穷时，出现问题的地方正是减法那一步。当时，我们发现在等式两边同时减去无穷时，等式是不成立的。所以如果我们能解释这个问题的话，这将会成为我们理解无穷可以是什么、不能是什么的一条非常重要的线索。

从数学的角度讲，我们可以说“每个数字都有一个相反数”。某数字加上其相反数可以把该数字带回到“原点”，也就是零。你怎么才能从1返回到0呢？你减去1。或者换一个说法，你加上–1。你怎么才能从2返回到0呢？你加上–2。你怎么才能从n返回到0呢？你加上–n。

要求所有的数字都有一个相反数，就好像要求你的每一个乐高小人儿都得有一个头盔一样。如果你是一个被宠坏的孩子，只要你提出这个要求，你的父母可能就会立刻为你所有的乐高小人儿订购头盔。数学的一大（并非广为人知的）乐趣就是你立刻就能获得你想要的一切，而不是非得有溺爱你的父母。一旦你要给每个数字设定一个相反数，嗖的一下，每个数字就都有相反数了。这让数学研究变得非常有趣，因为你能够获得任何你想得到的事物，唯一的附加说明就是你必须承担你的新玩具所带来的所有的逻辑后果。就好像如果你还是那个被宠坏的孩子，而你这次想要一只狮子，那么你的父母可能真的会给你找一只狮子来，但是你更可能会被狮子吃掉。类似的，如果你正在做数学研究，而你想让0等于1，那也没有关系，但是这样一来所有其他的数字也就都等于0了，你的数学世界就会崩塌，就像是被一只饥饿的数学狮子吃掉了一样。

如果你想让0等于2或者任何其他数字的话，这也都没有关系。但是这样你就陷入了一个叫作“模运算”的循环世界。设定0＝12是我们计算时间的方式，而设定0＝360是我们用圆规测量角度的方式。

无论如何，作为被宠坏的科学家，我们决定让每一个数字都有一个相反数。于是嗖的一声，每个数字都有一个相反数了。没有更多额外的事物，只是多了这些数字的相反数。自然数和它们的相反数一起构成了整数。从我们构建整数的方式中，我们知道整数包含下面这些内容：

- 0
- 任意自然数n
- –n，n代表任意自然数

使用更加直白的书写方式，我们获得了

…–4、–3、–2、–1、0、1、2、3、4…

从技术的角度讲，现在，减去4变成了加上–4。这听起来有一点儿复杂，但是对于数学家来说，这其实是让事情变简单了。因为我们不再需要定义减法这个新的运算类型，只用加法这个旧的运算类型就可以做所有的运算

了。

无穷属于这些新的数字吗？看起来不太像。但是我们能够确定这一点吗？可以。我们能用反证法证明这一点，就像我们用反证法证明了无穷不是自然数一样。具体而言，因为整数也遵守自然数的原则，也就是说，从等式两边减去同样的数，等式仍然成立。我们知道这对于无穷来说是不适用的，所以无穷不可能是一个整数。看来，这趟关于无穷不是什么的旅程还将继续下去。

事实上，从技术的角度讲，每当我们在发明一种新的数字类型时，我们都会获得一种“自由”。当我们发明自然数的时候，我们获得了加1的能力。我们把可以做加法的数学领域叫作可交换幺半群。自然数是一个自由可交换幺半群，对这类数字，我们可以自由地做加法。“自由”在这里是指没有附加的条件和限制，我们可以无穷次数地做数字间的相加。我们引入整数这个概念，是为了获得减法的能力。从技术的角度讲，我们是想要获得一个自由交换群。在数学领域里，自由交换群是指，在这里你可以自由地加和减。

为什么0没有倒数

我们用整数来推导无穷的尝试被证明是失败的，所以现在我们需要一条新的路径：不再尝试向上数到无穷，而是尝试把一个事物分成无穷份。你可能还记得在学校的时候曾经听老师说过，“你不能用数字除以0，否则的话就会得到无穷”。所以，也许我们恰恰可以通过这么做得到无穷——用一些数字除以0。也许我们可以定义一个分数，$\frac{1}{0}$？不幸的是，这好像也不行。我们最终会发现这种做法有其合理性，但真正的无穷并不以下面这个等式来表达。

$$\frac{1}{0} = \infty$$

事实比这个等式微妙得多。

想象一下你要把一个蛋糕分给零个人，那么每个人能够获得多少蛋糕？这个问题没有意义，因为实际上一个人也没有。你可以说，“每个人都得到了10个蛋糕！”因为本来就没有人，所以你也没有说错：这零个人每个获得了10个蛋糕。但是同时这零个人也可能每人获得了20个蛋糕，甚至可能每人获得了40个蛋糕和63个大象！因此，用1除以0显然不是有逻辑地推导无穷的方法。

既然如此，为什么老师要告诉我们“不能除以0”呢？在数学领域里，“我为什么不能……”通常不是一个正确的问题。正确的问题是，“我为什么可以？”从逻辑上讲，举证的责任是指你需要证明你做的所有事情是合理的。逻辑证明是数学世界搭建的基础。如果你不能通过逻辑证明一件事情，那么这件事情就不属于数学领域。你没有找到不做一件事情的理由，不代表你就有理由做这件事情。不过在现实生活中，我做一件事情既需要找到做这件事的理由，也需要找不到不做它的理由。我从事数学工作是因为我的爱好而不是因为数学的实用性。然而，如果数学真的没有用的话，那就意味着现在我有一个不做数学工作的理由了。

无论如何，为了更好地理解这个例子，我们需要先搞清楚除法的真正含义。除法绝对是＋、–、×、÷四种基本运算之中最难的一种。在学校里，除法也往往是最后学的一种运算。老师们在开始介绍除法时，常用的示例往往是分一些东西。之后老师们会告诉学生除法是乘法的“反向”，但一般并不会解释其中的理由。如果你从数字3开始做计算，3乘以4会得到12。为了重新回到3，你需要用12除以4（见图4–1）。

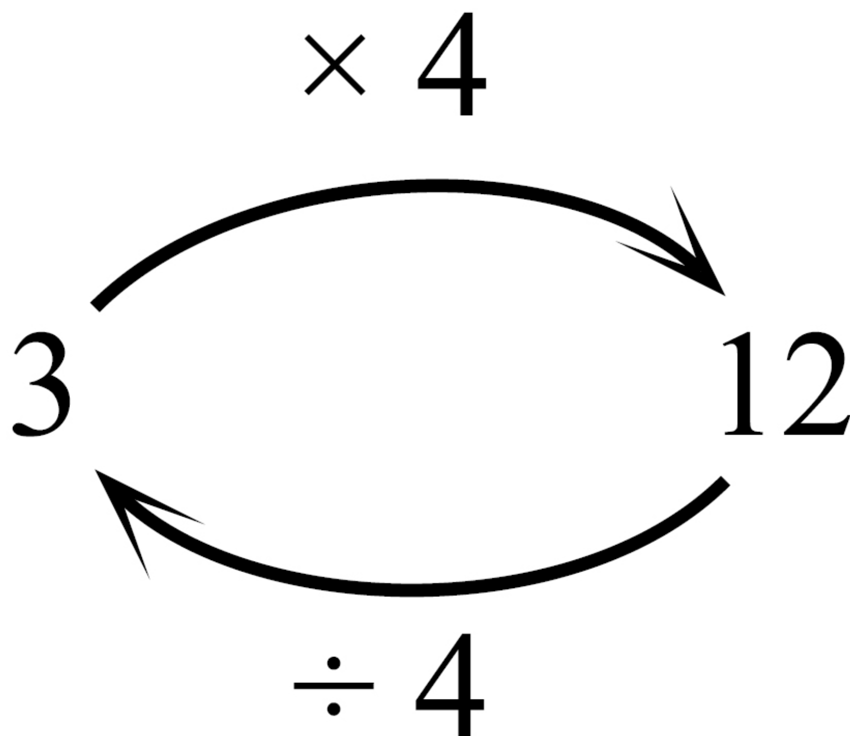


图4–1

这可能会让你想起我们为了“撤销”加法而创造了负数这个概念。确实如此，老师在介绍除法时采用的是同样的办法，只不过这次撤销的是乘法而不是加法。

在做加法时，我们的第一步是找到“什么都不做”的数字，也就是0。在乘法里也一样，我们需要找到和其他数字相乘而什么都不会发生的数字。这个数字就是1，因为：

$$1\times 2=2$$

$$1\times 3=3$$

$$1\times 4=4$$



$1\times n=n$，无论n是什么



所以，1被称作“乘法单元”，就像0被称为“加法单元”一样。

我们现在可以开始考虑倒数了。一个数的倒数可以把原来的数带回到1。

什么数乘以2能够返回到1呢？答案是$\frac{1}{2}$。什么数乘以3能够返回到1呢？答案是$\frac{1}{3}$。什么数乘以n能够返回到1呢？答案是$\frac{1}{n}$。

现在回忆一下，我们是怎样从自然数获得整数的：我们像一个被宠坏的孩子，希望为每一个自然数都找到一个相反数，于是我们就有了负数。也许你现在想要让每一个整数都有一个对应的倒数。你可以那么做，但是你要知道，可能会有一只数学狮子跑出来吃掉一切——因为想要得到0的倒数本身就是一个错误。

下面我会解释为什么你想要得到0的倒数是错误的。让我们把0的倒数这个数字称作x，因为我们还不知道它到底是什么。我们只知道，这个数乘以0会返回到1。换言之，

$$0\times x=1$$

但是等一等，$0\times x$应该是0才对。

事实上，这需要费一点儿力气才能证明。你可以说因为$0+0=0$，所以（$0+0$）$x=0x$。根据乘法分配律，我们知道等式左边是$0x+0x$，所以我们得到$0x+0x=0x$。现在，从等式两边都减去$0x$。左边得到$0x$，右边得到0，所以$0x=0$。

于是，上面的等式变成了：

$$0=1$$

天啊，这个不成立的等式又出现了。我们是不是就摆脱不了它了？整个推导过程教给我们的道理是，我们真的不应该想要得到0的倒数。因为如果0有倒数的话，我们最终就会得到$0=1$，然后所有的数字就都等于0了。

最后，还记得我们曾经说过减去一个数相当于加上这个数的相反数吗？同样的，除以一个数相当于“乘以这个数的倒数”。所以，除以2，事实上就是乘以$\frac{1}{2}$，除以3，事实上就是乘以$\frac{1}{3}$。那么，我们可以除以0吗？为了除以0，我们需要乘以0的倒数，但是我们刚刚才断定0没有倒数。反过来说，既然0没有倒数，那么我们当然不能除以0。

既然我们不能除以0，那么我们也就肯定不能把无穷定义成$\frac{1}{0}$了——因为$\frac{1}{0}$根本不存在，更确切地说，它根本不能存在，所以我们不能像想象负数那样想象它的存在。

构建有理数

所有的分数有一个总称，叫作有理数。把所有分数称作有理数并不是说它们可以坐在一起进行有逻辑的对话。它们是在整数的基础上，以整数与整数之比的方式构建出来的。

当我们使用自然数构建整数的时候，整个过程看起来很简单。我们只需要构建所有数字的相反数就行了。使用整数构建有理数比这稍稍复杂一点儿。第一步就是获得除了0以外的所有整数的倒数。但是这么做我们只能获得$\frac{1}{2}$、$\frac{1}{3}$、$\frac{1}{4}$这样的数字以及它们的负数版本。也即，这并不能帮助我们获得类似$\frac{4}{5}$这样的数字或者任何分子不是1的分数。为了获得这些数字，我们必须再一次像被宠坏的孩子一样提出要求：我们想要让一切能够相乘。在

构建整数的时候，我们并不需要提这样的要求，因为有了负数，我们就已经能够做所有的加法运算了。然而，获得所有的倒数并不会自动为我们带来所有乘法运算的结果。如果我们只构建出了倒数，我们只能获得$\frac{1}{n}$这样的数，其中n可以是任意整数。所以，总的来看，我们有了下面这些数字：

$$\cdots\ -4\quad -3\quad -2\quad -1\quad 0\quad 1\quad 2\quad 3\quad 4\ \cdots$$

$$\cdots\ -\frac{1}{4}、-\frac{1}{3}、-\frac{1}{2}、-\frac{1}{1}\qquad \frac{1}{1}、\frac{1}{2}、\frac{1}{3}、\frac{1}{4}\ \cdots$$

如果你挑出上面一排数字中的任意两个数并把二者相乘的话，你就会得到另外一个属于上面一排的数字。如果你挑出下面一排数字中的任意两个数并把二者相乘的话，你就会得到另外一个属于下面一排的数字。但是如果你把上面一排数字中的一个数和下面一排数字中的一个数相乘的话，那么你有可能会得到一个既不属于上面一排也不属于下面一排的数字。

所以，为了能够让一切数字相乘，我们需要获得所有上面一排数字和下面一排数字之间的乘积，这样我们就得到了所有的有理数。我们可以做出下面这样的总结：

有理数包括所有$\frac{a}{b}$这样的分数，其中

- a和b都是整数，
- b不为0，并且
- 如果ad＝bc，则$\frac{a}{b}=\frac{c}{d}$。

上面第三点是为了确保我们不会把$\frac{1}{2}$和$\frac{2}{4}$误认为两个不同的数。

在做了如此多的努力之后，我们还是没能得到无穷。我们可以用和之前一样的论证方式证明无穷不是一个有理数，因为有理数还是要遵守在等式两边同时减去同一个数后等式仍然成立的规则。构建有理数并没有帮助我们找到无穷是什么。数学的海岸线还在继续延伸。

无理数

无理的人是指非理性的人，而无理数是指不是有理数的数。你在学习无理数的时候，老师可能会说无理数是“小数位数有无穷位的数”或者是“一个数字的平方根”。这两种说法都包含着一些正确的理解，但是在严格意义上，二者讲得都不准确。

我们先看一下小数位数有无穷位这个说法。如果我们把$\frac{1}{9}$用小数的形式来表示的话，我们会得到0.1111111111…，这个小数的1会有无穷位。但是$\frac{1}{9}$肯定是一个有理数，而它的小数位数无穷多是因为我们武断地选择10作为小数基数，而10和9之间不是很好转换。10和5就很好转换，所以$\frac{1}{5}$的小数形式0.2看上去就非常干脆简洁。在这里，两个数字之间“好转换”的意思实际上是指二者“有公因子”。

如果我们用9作为基数展开$\frac{1}{9}$的话，我们会得到一个非常简洁的小数，0.1。

如果我们用3作为基数展开$\frac{1}{9}$的话，得到的则是同样非常简洁的0.01。

“小数位数有无穷位”这个说法背后的真相是无理数的小数部分无穷延续而且不循环。这个问题在大学的数字分析课上证明起来非常有趣。但是用这种方法来检查一个数是不是有理数显然是非常令人崩溃的。要把一个数字的小数位数扩展到一百万位，一个人即使不眠不休也要花上一个星期。但是即便你这么做了，你还是不知道这个小数会不会在小数点后第二百万位之后开始循环，会不会在小数点后第一亿位或第一万亿位之后开始循环。

“某个数的平方根”这个说法也是有问题的。因为4的平方根是2，而2显然是一个有理数。而且还有π和e这样的无理数，这些无理数不是任何有用的数字的平方根。当然，我知道π是π2的平方根，但是这并不能帮我确定π到底是不是无理数。这就是为什么唯一能够让人接受的“无理数”的定义就

是“不能表示为两个整数的比值的数”。

你可能会反驳说$\pi = \frac{22}{7}$。但是这个著名的分数仅仅是π的一个近似值，是人们还没有计算器的时候用来简化计算的。而现在手机上的计算器都已经有π这个按钮了。π和$\frac{22}{7}$只有个位和小数点后前两位是一样的，也就是3.14。在日常计算中，这个精确度就已经足够了，比如蛋糕配方涉及的计算就不需要太高的精度。但是从数学的意义上来讲，这两个数字完全不同。（事实上，在计算蛋糕配方的时候，用3作为π的近似值已经足够精确了。）

现在，你也许会觉得几乎所有的数都可以被看成两个整数的比值，只不过对于像$\sqrt{2}$这样的数来说，论证过程会比较简单；而对于像π和e这样的数，证明起来会复杂得多。

无论如何，这些都不能帮助我们定义什么是无理数。我们不能说无理数是“所有不能写成两个整数的比值的事物”，因为那意味着大象也是无理数。但是我们也不能说无理数是“所有不能写成两个整数的比值的数”，因为什么是数呢？我刚刚兴高采烈地举了e的例子，但什么是e呢？f、g和h也都是数吗？如果π是一个数的话，α是不是也是一个数呢？

构建无理数非常困难，所以我决定我们在这本书后面一点儿的地方再讨论这个问题——等到讨论无穷小的事物的时候再讨论。不过其基本逻辑还是和之前一样。我们还是要像被宠坏的孩子一样要求获得做一些事情的能力，然后我们就获取了做这些事情所需要的工具。问题在于，这次我们想做的事情是什么？答案大概是：“填充有理数之间的空白区域”。事实上两个有理数之间总会有一些间隔。我们可以用下面这样的一条线上的一些点来表示整数，也就是整数轴（见图4–2）。

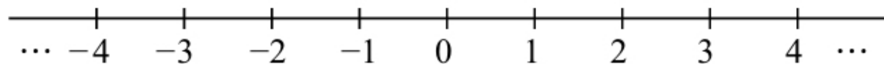


图4–2

我们知道，这些整数之间存在很多有理数（见图4–3）。

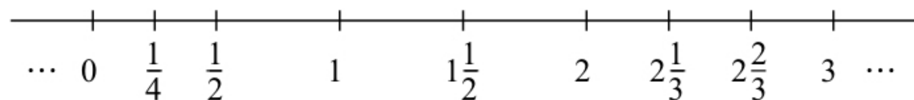


图4–3

但是，即便我们把所有的有理数都填上，这些数字之间还是有间隔。虽然现在我还没有解释我们怎样填充这些间隔，但下面是一些我们在做这件事的过程中发现的有趣的事实：

- 任何两个有理数之间都存在一个无理数。

- 任何两个无理数之间都存在一个有理数。

这听起来很像奇数和偶数之间的关系，因为奇数和偶数总是相互交叉地排列在整数数轴上。除此之外，下面还有一个和我们的直觉不相符的事实：

无理数的个数比有理数多。

这怎么可能？如果无理数与有理数交替存在，那么怎么可能其中一个比另一个的个数多呢？这是另外一个和无穷有关的神秘事实。在这里，我们考虑的不仅仅是无穷个数字，还要考虑无穷接近彼此的事物。对于奇数和偶数而言，我们可以选择任意一个偶数，然后轻易推算出与它相邻的奇数。比如与偶数6相邻的奇数是5和7。然而，在有理数和无理数的世界里，它们彼此之间是如此接近，以至于我们无法说出那个和某个有理数相邻的无理数是什么。无论我们说出哪个数，总是存在比它更加接近那个有理数的无理数。

以数字1为例。比1大但是又最接近1的有理数是什么？并不存在这么一个数。如果你说是数字x的话，那么我总是能够找到数字$\frac{1+x}{2}$比x更接近1。这就相当于取1和x的平均数。这个数刚好在1和x中间，比1大而又比x小。而且这个数一定是有理数。因为如果x是有理数，那么1＋x就是有理数，其除以2之后还是有理数。你可以一直对分从1到x的这段距离，同时越来越接近1。你得到的有理数会越来越小，但是仍旧大于1。就像吃一块巧克力，你每次都只吃剩下的一半，那么这个过程就可以一直持续下去。你永远也无法真正到达1，在你说出的数字和1之间总还有其他的有理数。

你可以把这件事想象成把这条数轴不停地放大比例，或者你可以在你的计算机上直接这么做。如果你以数字6为中心放大一条整数轴，你总会到达这样一个放大倍数，此时你的屏幕上只剩下6这个数，而所有其他的数字都掉到屏幕外面去了。但是如果你放大一个全部空隙都被填满了的实数轴，那么无论你放大多少倍，屏幕上总还是会有更多的有理数和更多的无理数。所以你不能说有理数和无理数是“交替”出现的。事实比那奇怪得多。

有理数和无理数的总称是实数。而上面这种能够一直放大的属性叫作有理数和无理数在实数中的密度。有理数不是实数的全部，但是有理数在实数中的密度是如此之大，以至于无论你放大多少倍，你还是会看见有理数。无理数也是一样。我们将在第15章继续讨论这个问题，到那个时候，我们会进一步理解无穷小的距离是什么。

那么，这些讨论能帮助我们弄懂无穷到底是什么吗？不能。我们还是可以证明，在实数的范围内，在等式两边同时减去一个数，等式仍然成立。而如果无穷是一个实数，那么我们还是会推导出$0=1$这个令人讨厌的悖论。所以，无穷不是一个实数。

所以，现在，你可能已经开始意识到探寻无穷不是什么的努力完全是徒劳，除非我们能创造出一种数字来。在这种数的范畴里，在等式两边同时减去一个数后等式不成立。这种数字会是什么样子的呢？到目前为止，我们每次发现一种新的数字类型都是以要求获得一种新的能力为前提的，比如减法、除法和填充空隙。每次我们要求获得一种新的能力的时候，我们都能发现一个新的世界，找到一种新的建造数学大厦的材料。这就像你现在想要找到一种新的烹饪配方，所以你认为自己需要一个有更多新食材的厨房。

奇怪的是，我们真正想要的是这样一个世界，在其中，我们能做的事情更少而非更多，因为我们想要找到的是一个等式两边同时减去一个数后等式不成立的世界。有时候当我准备出门旅行的时候，我会计划非常多不同的活动，结果就是，我发现自己往行李箱里放了越来越多的用于不同场合的东西。因为我不希望自己成为那样一种数学家，他们无论出席哪种场合都穿相同的衣服——不管是演讲、徒步旅行、音乐会，还是沙滩聚会。然而最终我扔进了如此多的鞋子，以至于我完全无法提起我的行李箱。换言之，我已经没有办法出门了。现在，我们在数字方面也已经到了这样的一个阶段：我们需要把所有的东西都扔出去，然后重新开始。

---

1. 火焰雪山是一道非常有意思的甜点，它最里面的芯是一块冰激凌，中间是一层海绵蛋糕，最外面是一层蛋白泡沫。——译者注

5　一个函数问题

一次，我像往常一样走路回家，途中我发现了一家新开的服装店。因为我从来没看见过附近进行过施工，所以我很好奇地问他们是什么时候开业的。他们说，十年前。

我不知道为什么我直到那一天才突然注意到这家店，而在之前上千次路过这里的时候都没有注意到。有时候走得慢一点儿或者稍稍换一个视角，我们就很可能会在自己觉得非常熟悉的地方发现新鲜的事物。现在，我们要稍稍变换一下看数字的视角，也许这么做就会让无穷自己出现在我们面前。

我们曾经试图数到无穷，但是没有成功。也许你无法相信，那可能是因为我们数的方式太高级了。孩子们在最初学习数数的时候，并不是用不断加1的方式数数，他们数的是自己的手指。

我敢说，很多小孩子都和我一样，小时候最喜欢的吃呼啦圈糖（Hula-Hoops）的办法就是把它们套在手指上。其他一些地方的读者可能没有听说过这种环形的、用土豆做的小零食，但是有人告诉我他们那里的小孩子会用同样的办法吃妙脆角。虽然妙脆角是锥形的而不是环形的，但是吃的方式也没什么差别：你把呼啦圈糖或者妙脆角套在每一个手指上，像魔术师一样甩一甩手指，然后就可以开始一个一个地吃掉它们了。

扳着手指数数通常被看作一种低级的数数方式，只有还没有学会在脑子里数数的小孩子才会用这种办法。我倒是想提出一种不同的观点。扳着手指数数其实是一种意义深远的办法，它更有效率，而且能帮助我们到达无穷。至少，这种数数的思路能把我们带到无穷，虽然我们的手指本身不能到达无穷。

扳着手指数数的一个精妙之处在于你不需要费力记住你数到哪里了，你甚至不需要按照顺序数。你只需要让每一个手指指代你要数的一个事物，如果你要数的东西有10个，那么手指用完了，你也就数完了。至少对于我来说，这种办法的高效之处在于帮助我们省下了记住当前数到哪个数字的脑力。我们可以用省下来的脑力思考一些其他的事情。比如说，我不用费力记住我现在数到4了，因为我的手指用到了第4个，与此同时我就可以在头脑里思考一些更加复杂的事情了。你可能会觉得我有点儿不可理喻——毕竟，我是一个数学家，我怎么能讲这种话呢？我是不是在取笑数学家数不好数？

下面是一个常常出现的情形。有一个朋友要过来喝咖啡，我想把4勺咖啡粉放到咖啡机里——我喜欢每人两勺咖啡粉的浓度。同时出于礼貌，我想继续和朋友的对话。通常来说，我会在数到第2或第3勺的时候出错。我可能会因为忘了自己已经放了2勺还是3勺而不得不停止与朋友的对话。很明显，谈话所占用的脑力限制了数数可占用的脑力。那换一个方法呢？如果不是在脑子里数数，而是用手指数数，我一般就不会出错了。当然，一些不像我这么讲礼貌的人可能就会直接中止与朋友的对话；不像我这么对数学斤斤计较的人可能也不在乎多放或者少放了一勺咖啡粉。但是谁让我这么讲礼貌又这么在乎数学的精确性呢！

说完了精确的部分，下面我来说说为什么我觉得扳着手指数数是一种意义深远的方法。首先，从数学的角度定义一下10。什么是10？早些时候，我们曾经介绍过，10就是：

$$1+1+1+1+1+1+1+1+1+1$$

但是，如果你之前并不知道“10”这个概念的话，你怎么去找到10呢？

10就是我们的手指的总数。任何事物，如果其数量等于我们手指的个数，那么我们就说其数量为10。这听起来不太像数学家说的话，是不是？事实上，当数学家第一次在严格意义上定义数字的时候，他们基本上就是这么做的——虽然不一定是以手指的个数为标准。

所以，我们知道了怎么数10以下的事物的数量。一个朋友给我看了她的孩子的家庭作业，里面有很多关于“实际生活”的问题，比如在“实际生活”中数一数小点心的数量。假设我们也面临这个问题，我们要怎么做呢？

首先，我们已经认定10就是“我们手指的数量”。然后，我们给每一个手指取一个名字，这样在计数的时候，除了扳手指之外，我们还可以在脑子里叫它们的名字，或者干脆直接把它们的名字喊出来。我们可以管我们的手指叫汤姆、史蒂夫、皮耶特、尼克、理查德、艾米丽、多米尼克、约翰、尼尔和艾丽莎，或者我们也可以管它们叫1、2、3、4、5、6、7、8、9和10。然后我们把这些名字对应到每个小点心上。现在，每个小点心都有了对应的名字，而通过念出这些名字，我们知道这些点心对应上了我们的手指。我用图5–1来总结了我刚刚说的过程。

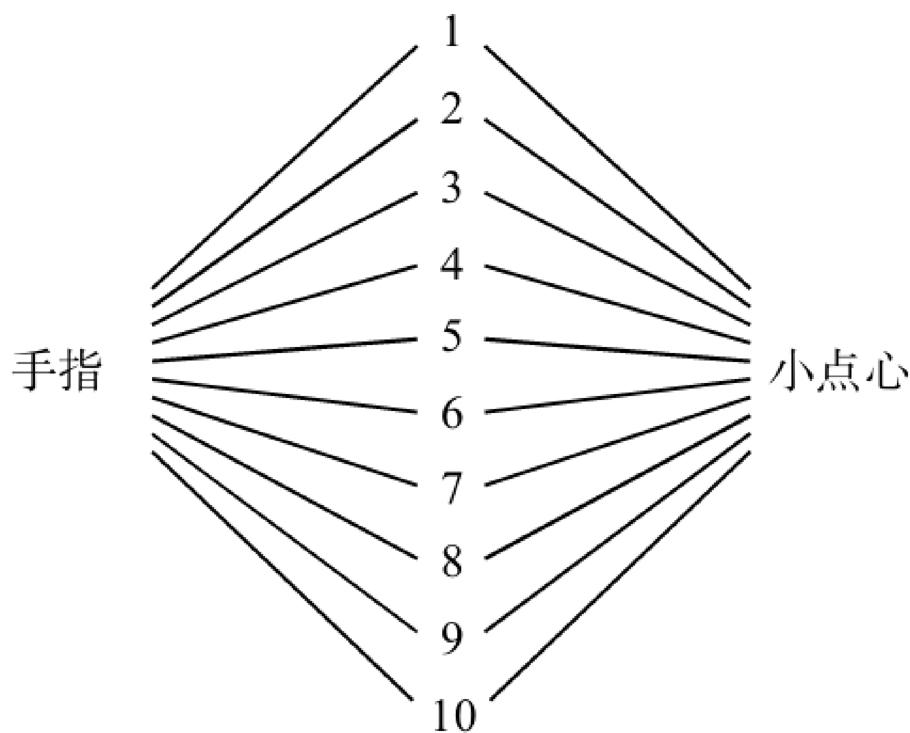


图5–1

从图5–1可以看到，我们事实上在中间增加了一个步骤。我们没有直接把点心和我们的手指对应起来，而是先把我们的手指和一些名字对应起来，然后把这些名字和点心对应起来。

当然，我们这么做的原因是我们要数的事物经常会超过10。为了避免我们没有手指可用，我们就采用了这样的办法。你也可以试着用上你的脚趾，或者把你的手指当作二进制的位数使用，这样你就能用手指数到1023了。对于后一种方式，我们会在第7章详细介绍。

然而，尤其是对于我和我容易被谈话打扰的头脑来说，跳过中间的步骤有时更讨巧而且有效。假设你在一个餐馆订了一个16人的位置，你的朋友们三三两两地聚在桌子周围谈话。因为每个人都在动来动去，你可能很难搞清楚到底到了多少人。而如果让每个人都坐下来的话，我们就能很容易地看出来谁还没到。你并不需要真的去数人的数量，而只需要检查是不是每个人都和他的座位对应上了。

这是数数的数学版本。你可能会觉得这个说法很奇怪：数数本身不就是数学吗？我这么说的意思是，这是用数学的方式让计数这件事变得更加严谨。

数学的严谨性

现在我想稍稍跑题一会儿，简单聊一聊数学的严谨性。数学的严谨性是数学家们能够针对某个论断是对是错达成一致的基础。只有在这个基础上，数学家们才不会在理论观点方面各执一词，无休无止地争论下去。数学是建构在逻辑原则上的。如果你的研究对象严格遵守逻辑原则，而你的操作也严格按照逻辑原则实施了，那么你得出的结论别人就不会提出异议。

然而，如果你的研究对象像人或者云一样，并不严格遵守逻辑原则，那么不同的人就可能会得出不同的结论。同样的，如果你的操作没有严格按照逻辑原则实施，结论也可能会有所不同。总体上讲，我们所处的世界并不严格遵守逻辑原则。如果你给一个孩子一块饼干，然后又给他另一块饼干，很可能最后孩子手里不是两块饼干而是一块饼干也没有——除非你算上他肚子里的饼干。

数学的起点就是消除模棱两可，只留下能够明确地用逻辑来处理的事物。数学家们的工作就是使用逻辑来观察这些事物的变化。无论是在你自己的生活里，还是在数学的历史上，接下来发生的事情都会让人觉得有点儿纠结，因为一些看起来非常“显而易见”的事情会变得非常难以理解。那么，这么做的意义何在呢？

这么做的意义之一就是理解那些不那么显而易见的事物。如果一件事在你的直觉里并不那么显而易见的话，那么我们就需要找到理解它的路径。无穷就是一个例子。我们好像找不到能够合理地处理无穷的方法。我们到目前为止付出的所有努力都得到了“$0=1$”这个错误的结论，也就是说，这些努力都没有结果。我们接下来要做的就是重新思考我们处理有限的数字的方法，从中寻找能够帮助我们理解无穷的途径。我们将要重新思考我们认为最明显也最基本的数字，也就是用来计数的数字。

用袋子数数

从本质上来讲，计数就是把一组事物对应到另外一组“官方”的事物上，而这些“官方”的事物就是我们所说的数字。我们已经知道了怎样用我们的手指定义10。我们的手指对于10来说，就是官方的标准物。每当有一组事物能够刚好对应到我们所有的手指上的时候，我们就知道这组事物的数量是10个。

我们可以为每一个数字指定一个官方标准物。比如，我们可以将一个袋子里的东西的数量作为23的官方标准物，这样一来，23就被定义成了“那个袋子里的事物的数量”。这样，每当人们想要数23个东西的时候，他们就会把要数的东西和那个袋子里的东西比一比。这听起来非常傻，但还是有一定道理的。回想一下我们是怎么定义“千克”的：巴黎附近的一个金库里面有一块官方认定的金属块，这个金属块的重量就是1千克在严格意义上的定义。

数学家并没有用在物理上存在的官方的袋子来定义数字，因为数学是一个抽象的学科而不是一个关于实际物质的学科。所以数学使用了官方的“抽象的”袋子来定义数字。这些袋子是抽象的，因为它们看不见摸不着。但它们在理论上是存在的，而且数学家们并不把它们称作袋子，而是称作数集。不过，这不妨碍我们把这些数集想象成袋子。

我们需要想象的第一个袋子，它的里面装着零个东西。换言之，这个袋子是一个空的袋子。我们把零定义为“这个袋子里所装的东西的数量”。也许我们可以给这个袋子加一个编号，写上0（见图5–2）。



图5–2

现在，我们需要制作一个包含一个物体的袋子。四处看一看，我们能放进去一点儿什么呢？现在，我们所拥有的一切就是一个空袋子。不过它也完全符合要求。所以我们就把这个空袋子放进第二个袋子里。这样，第二个

袋子就定义了数字1。然后我们给这个袋子编号为1（见图5–3）。

现在，我们需要做一个包含两个物体的袋子。而我们面前恰好有两个物体：0号袋子和1号袋子。所以理所当然的，我们往第3个袋子里放入这两个袋子，它的编号就是2（见图5–4）。

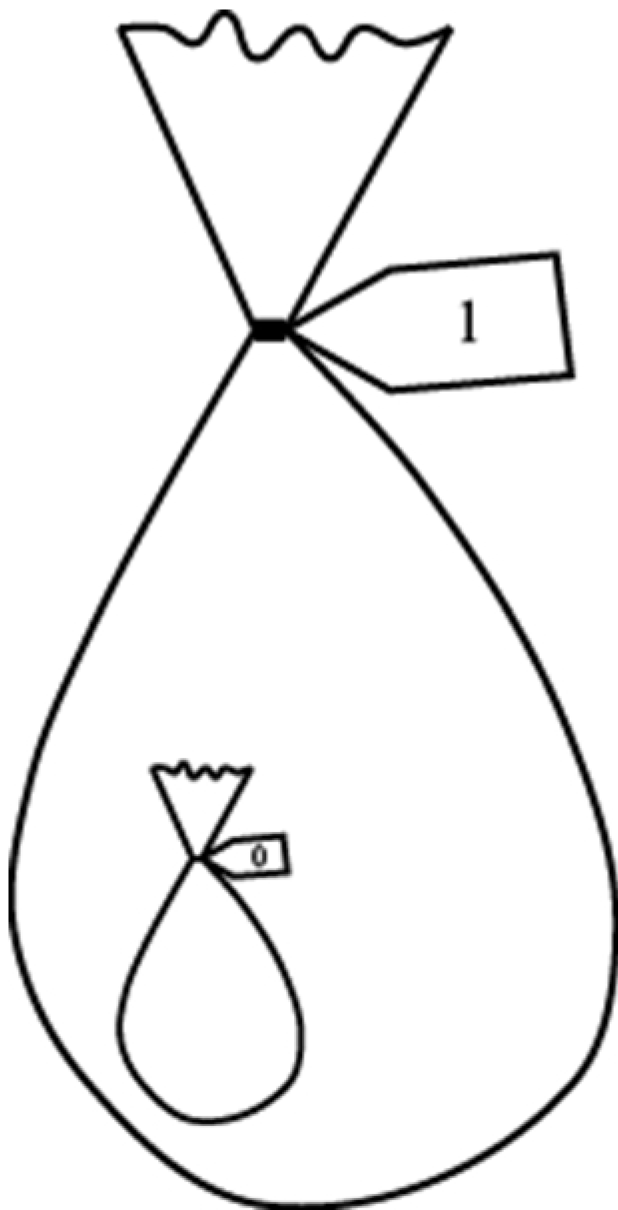


图5–3

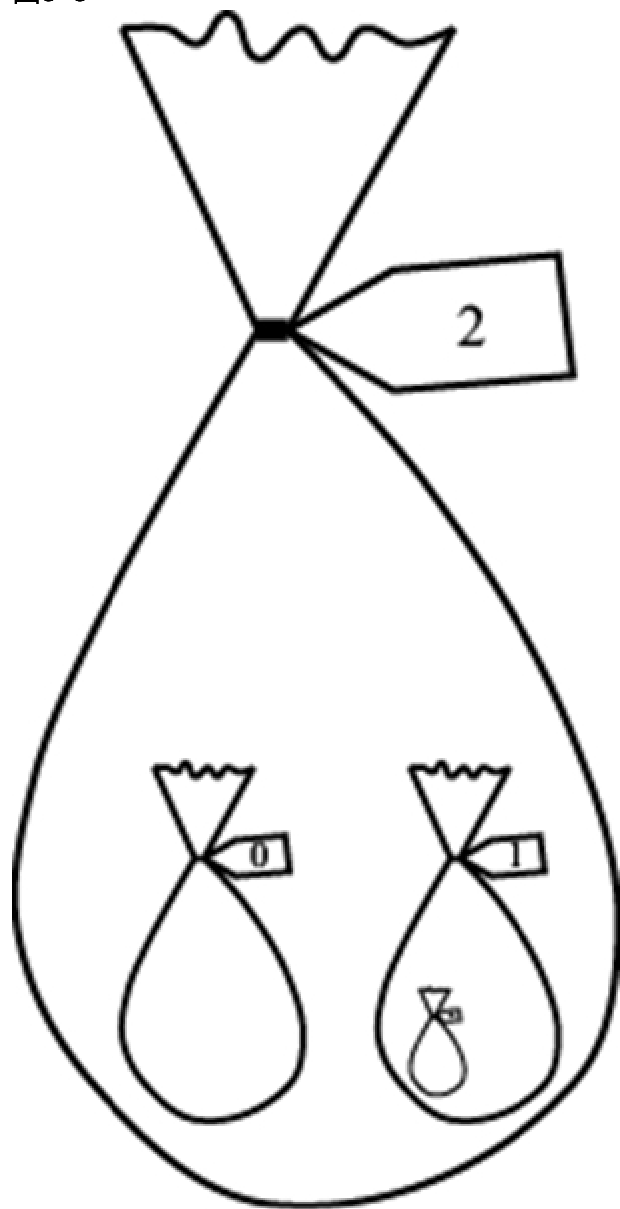


图5–4

现在，你可能会觉得不太对劲。你可能会抗议说2号袋子里面不是有三个东西吗？但是，你不应该去看里面的袋子。你应该把2号袋子里的两个袋

子当作两个整体，忽略它们里面可能会装着的东西。就像你打开一个装着麦丽素（Maltesers）糖果袋的袋子，你可以在不计较每个糖果袋里有多少颗麦丽素糖的情况下直接数大袋子里有多少个麦丽素糖果袋。

重复之前的步骤，我们还能得到3号袋子（见图5–5）。

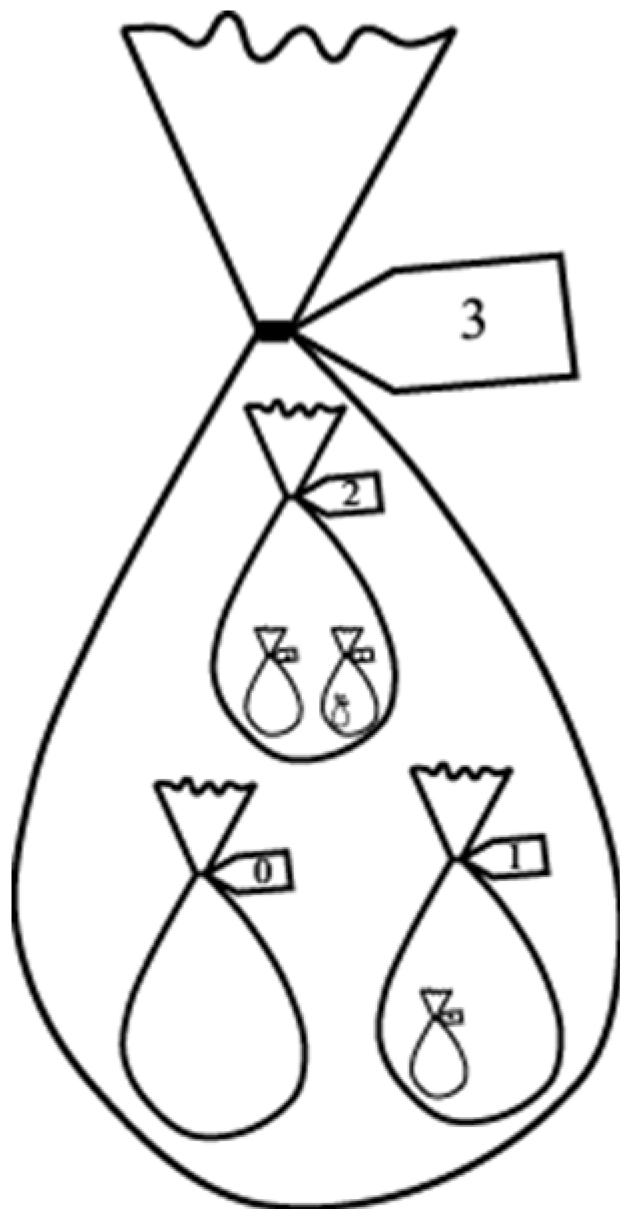


图5–5

说到麦丽素，你可能会有疑问，为什么我们不在袋子里装一些更有意义的东西，比如鹅卵石或者计数筹码？毕竟，计数筹码的唯一用处就是计数嘛。答案在于，虽然数这些物体看起来更有道理，但是这些物体在抽象的数学世界里是不存在的。袋子或者说数集是我们现在能利用的唯一的起始点，我们需要利用这些概念来构建整个数学世界。这种方法被称作集合论。当然，也存在其他构建数学基础的理论，但是集合论这个方法能够帮助我们最终定义无穷。

一旦我们获得了装有0个物体、1个物体、2个物体、3个物体……的“官方”袋子，我们就可以开始数其他的袋子里有多少物体了，方法就是：把新的袋子里的物体数量和官方的袋子做对比。我们必须小心翼翼地对比：

- 新的袋子里的每一个物体都要与官方袋子中的物体对应上。
- 不能把新袋子里的两个物体对应到官方袋子里的一个物体上。
- 你需要用尽官方袋子里的每一个物体。

这就像是在《舞动奇迹》节目里把专业选手和挑战明星一一配对一样：

- 每一个明星都必须和一个专业选手搭配。
- 不能有两个明星和同一个专业选手搭配。
- 每一个专业选手都要有人与之搭配。

结果是，我们会有同等数量的明星和专业选手。按照一一配对的原则，如果其中两个明星不得不和同一个专业选手搭配的话，明星的数量就比专业选手多了。如果不是每一个专业选手都有人与之搭配的话，那么专业选手的数量就比明星多了。但你也可以扔掉一一配对的原则。你可以有同样数量的明星和专业选手，然后安排其中的两个明星和同一个专业选手搭配，剩下一个专业选手无所事事。用列表的形式来表达可能看起来更清楚一些（见图5–6）。

从图5–6中我们可以看到，每个明星（左列）都刚好有一个专业选手（右列）与之搭配，但是专业选手特伦特·威顿没有与其搭配的队友。所以我们

不用数两者的数量就可以得出结论：专业选手的数量比明星的数量多。

托姆·伊文思	→	伊韦塔·卢库斯沃特
卡洛琳·福莱士	→	特里斯坦·麦克马努斯
艾丽森·哈蒙德	→	阿利亚·斯可雅尼克
史葛·米尔斯	→	乔安妮·克利夫顿
瑞秋·史蒂文	→	凯文·克利夫顿
西蒙·韦伯	→	克里斯蒂娜·瑞哈诺夫
马克·赖特	→	凯伦·豪尔
		特伦特·威顿

图5–6

然而，如果其中多个明星都和凯伦·豪尔搭配的话，在特伦特·威顿还是没有与其搭配的队友的前提下，我们就没有办法在不数两者数量的情况下确定明星和专业选手哪个数量更多了（见图5–7）。

托姆·伊文思	→	伊韦塔·卢库斯沃特
卡洛琳·福莱士	→	特里斯坦·麦克马努斯
艾丽森·哈蒙德	→	阿利亚·斯可雅尼克
史葛·米尔斯	→	乔安妮·克利夫顿
瑞秋·史蒂文		凯文·克利夫顿
西蒙·韦伯		克里斯蒂娜·瑞哈诺夫
马克·赖特		凯伦·豪尔
		特伦特·威顿

图5–7

将两组事物配对在数学里被称为函数。假设你有两个物体集合，我们把它们称为C和P。以C为自变量，P为因变量的函数将集合C中的物体与集合P中的物体对应起来。这样就满足了前面说的第一个条件，也就是C集合中的每一个物体都有且只有一个P集合中的物体与之对应。

前文说的第二个条件是只有一部分特定的函数才具有的特性。“不能有两个明星和同一个专业选手搭配”这类条件被称作内射性。而第三个条件“每一个专业选手都要有人与之搭配”被称为满射性。当我们找到一个同时满足三个条件的函数的时候，我们就可以称该函数为完美映射。这时候，其中一个集合中的每一个数字都会在另一个集合中找到唯一一个数字与之对应。没有人有多于一个的搭配队友，也没有人被剩下，如图5–8所示。

托姆·伊文思 → 伊韦塔·卢库斯沃特
卡洛琳·福莱士 → 特里斯坦·麦克马努斯
艾丽森·哈蒙德 → 阿利亚·斯可雅尼克
史葛·米尔斯 → 乔安妮·克利夫顿
瑞秋·史蒂文 → 凯文·克利夫顿
西蒙·韦伯 → 克里斯蒂娜·瑞哈诺夫
马克·赖特 → 凯伦·豪尔

图5–8

这种一一对应的搭配在数学中被称为双射函数或者双射性。

你可能会奇怪，为什么我们不再增加另外一些条件，比如“不能有两个专业选手和同一个明星搭配”。这是因为类似的条件已经自然地包含在第一个条件中了。而第一个条件就是函数的基本定义。明星和专业选手在《舞动奇迹》中分别扮演不同的角色，他们相互搭配，这和函数计算的原理是一样的。另一个思考方式就是，想象一台自动售卖机：左边的事物就是你在选择商品的时候需要按的按钮或者输入的代码，而右边的事物就是你可以买的商品。有可能出现不同的代码代表同样的商品的情况，比如，确实有一些自动售卖机，它每一个位置上放的都是听装可乐。但是，如果你输入同一个代码却得到了不同的产品，那么这台售卖机就有问题了。或者说，这样的自动售卖机可能会让你抓狂，因为你无法预测自己将会得到的

产品是什么。也许真的会有这样的自动售卖机吧，但它肯定不是我们要讨论的那种。

这种对应关系将会带给我们第一个关于无穷的有效定义。在此之前，让我们先来看一看这个理论在其他的数学情形下是怎么运作的。用专业术语讨论函数会非常枯燥无趣，但是我们可以用一些示意图来表示配对关系。示意图能够帮助我们搞清楚正在讨论的问题，但是对于数学家来说，使用干巴巴的专业术语更加合适，它可以确保其所做的数学工作不产生歧义，也不容易出错。不幸的是，在现实世界中，数学教学往往只剩下干巴巴的术语而丢失了直观的感受。

下面类比了《舞动奇迹》中明星与专业选手的配对方式，不过这里我们用数字代替了人名（见图5–9）。

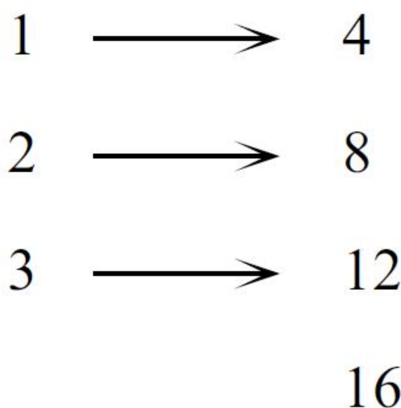


图5–9

更加精确地说：C是包含数字1、2、3的集合，P是包含数字4、8、12、16的集合。

这个函数为C集合中的每一个数字都搭配了一个是其4倍的数字。这样，1就对应到4，2就对应到8，3就对应到12。而16则被剩在一边，没有发生任何对应。这个函数并不满足前文所说的第三个条件，也就是说，它不是满射函数。因为16并没有与之对应的C集合的数字，就像图5–6中的特伦特·威顿一样。

✱下面是两个明星对应同一个专业选手的数字示意图（见图5–10）。

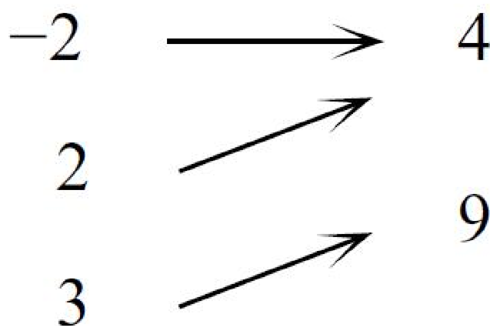


图5–10

C是包含数字–2、2、3的集合。

P是包含数字4和9的集合。

这个函数为C集合中的每一个数字都搭配了其平方数。这样，–2就对应到4，2也对应到4，3就对应到9。这里，–2和2对应到了P集合中的同一个数字。所以，这个函数并不满足前文所说的第二个条件，它不是内射函数。

✱下面是明星和专业选手完美搭配在一起的数字示意图（见图5–11）。

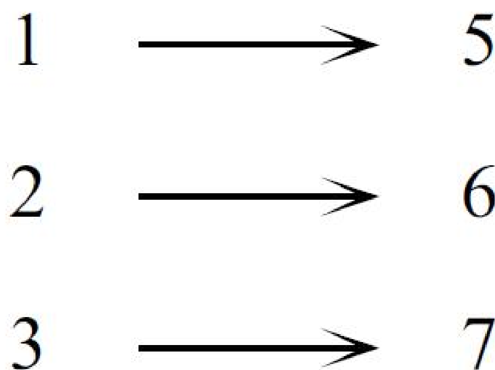


图5–11

C是包含数字1、2、3的集合。

P是包含数字5、6、7的集合。

这个函数为C集合中的每一个数字都搭配了一个比它大4的数字。这样，1就对应到5，2就对应到6，3就对应到7。我们可以看到，不存在两个箭头指向同一个数字的情况，右边也没有被剩下来的数字。这就是集合C和集合P的完美映射。它意味着左边和右边有同样多的数字，就像我们能把小点心和我们的手指对应起来一样。

现在，让我们大胆尝试一下无穷数集的对应（见图5–12）。

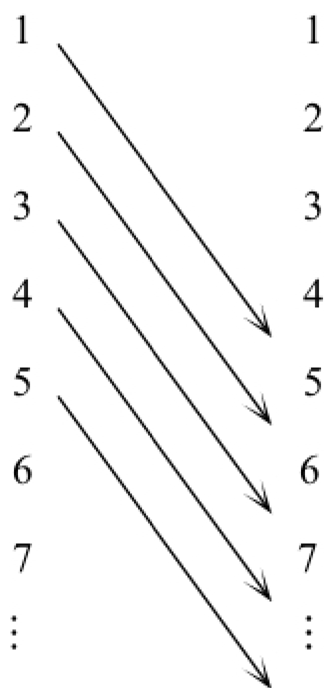


图5–12

C是包含所有自然数的集合。

P也是包含所有自然数的集合。

就像上一个例子里的函数一样，这个函数为C集合中的每一个数字都搭配了一个比它大4的数字。同样的事实还有，并不存在C集合中的两个数字对应P集合中的同一个数字的情况。但是这一次，P集合中的一些数字没有C集合中的数字与之对应：数字1、2、3、4都属于这种情况。所以这个函数是内射函数，但不是满射函数。

在下面这个例子里，我们把所有的自然数和所有的偶数对应起来（见图5–13）。这有点儿像我们在讨论希尔伯特旅馆问题时所做的事：把每个客人的房间号乘以2。

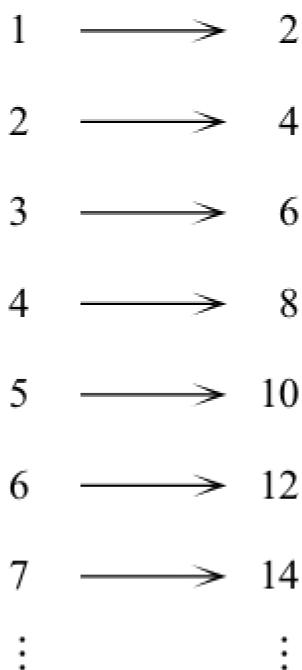


图5–13

C是包含所有自然数的集合。

P是包含所有偶数自然数的集合。

这个函数为C集合中的每一个数字都搭配了一个是其两倍的数字，没有两个箭头指向同一个数字，而且P集合中的每一个数字都有C集合中的数字与之对应。所以这是一个完美映射。

*下面这个例子有点儿像那个要把两层的希尔伯特旅馆中的客人转移到一层的旅馆中去的问题。我们需要把“明星”一侧的红色数字和蓝色数字，对应到“专业选手”一侧的紫色数字上（见图5–14）。

C现在是包含两个自然数集的集合，既包含红色的自然数，也包含蓝色的自然数。

P是紫色的自然数的集合。

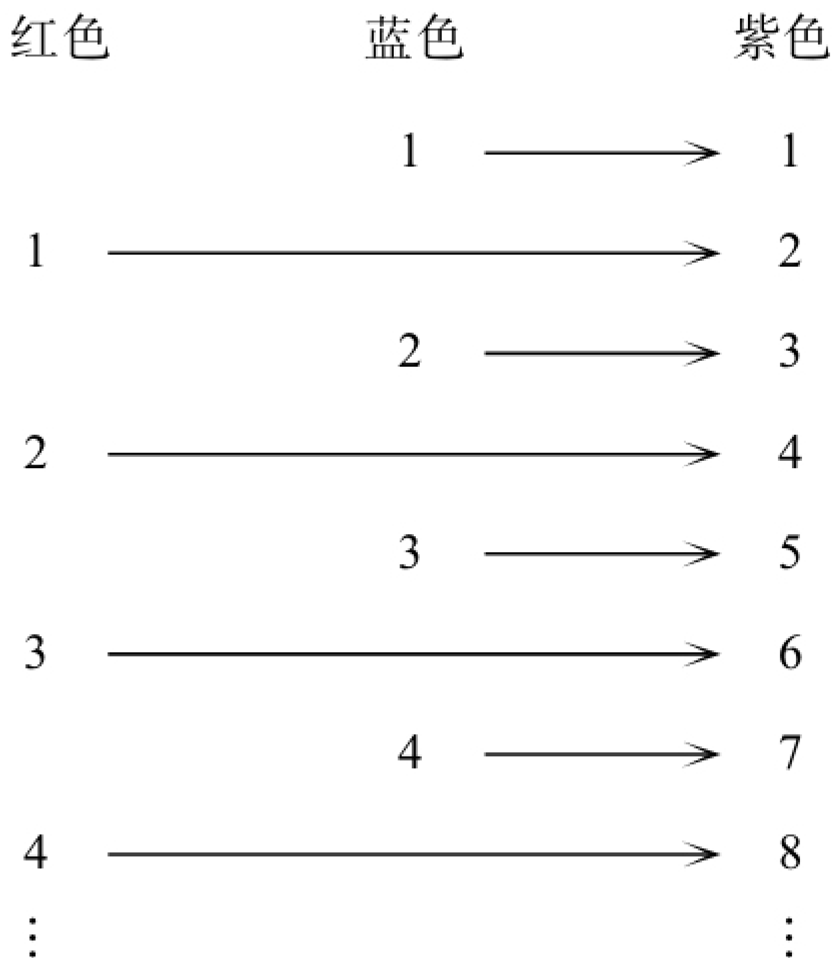


图5–14

这个函数的工作原理和转移两层的希尔伯特旅馆的客人是一样的。红色的数字n对应于紫色的数字2^n，而蓝色的数字n对应于紫色的数字$2^n–1$。从图5–14中我们能够看到，这个函数同时满足内射性和满射性的条件。不存在两个箭头指向同一个紫色数字的情况，而且所有的紫色数字都用上了。所以这个函数也是一个完美映射。

上述几张图示展示了如何构建包含无穷数集的函数，而同时让函数满足完美映射的条件。前面我们说过，如果你能够精确地把挑战明星和专业选手

搭配起来，即便不一个一个地数，你也知道你有同样多的挑战明星和专业选手。也就是说，如果你能够在两个数集之间搭建双射函数，就说明这两个数集具有同样多的数字。这个办法的聪明之处在于，因为你不用一个一个地数就可以比较数集的数量，所以你可以在无穷数集中使用这个方法。如此，这个办法就成为我们解锁无穷的钥匙，也就是数学领域中“可数性”这个概念。

可数性

学着数数听起来像是小孩子做的事。确实如此，每个人从很小的时候就已经开始学习如何数数了。但是这不意味着每样事物从此都可以被我们轻易地数清楚。在现实生活中，有很多东西在很多层面上是很难数的：可能是因为它们移动得非常快，比如操场上的孩子和田野里的兔子；可能是因为它们看起来非常像，所以你不知道哪个数过了，哪个没数过，比如树上的叶子；可能是因为它们非常小，比如沙粒；也可能是因为它们太多了。你数过的最大的数字是多少？我不认为我曾经数到过200以上。在我常常失眠的那个时期，我确实数过绵羊，因为反正别的办法也没有什么效果，为什么不试一试呢？但是我仍然不觉得我数到过200——还没睡着，我就已经数烦了。

在日常生活中，我们肯定不会数到无穷，倒不是说我们数烦了，而是说我们绝对不可能到达无穷的终点。但是在数学里，数数并不意味着要大声喊出“1、2、3、4……”数数意味着将你要数的物体和官方的计数袋子相对照。下面是这个问题的关键所在：

* 对应1的官方袋子里有1个物体：0号袋子。

* 对应2的官方袋子里有2个物体：0号袋子和1号袋子。

* 对应3的官方袋子里有3个物体：0号袋子、1号袋子和2号袋子。

现在，我们知道了怎么制作n号袋子，从而我们也就知道了怎么制作n＋1号袋子：把0号袋子、1号袋子、2号袋子……n号袋子全都装进这个新的官方袋子里。

我们在这里所做的事情是重新定义了从一个数字数到另一个数字的途径，也就是更新了“加1”的方式。我们把这个新的方式描述为“把前面所有的袋子都装到一个新的更大的袋子里”。戏剧的高潮一幕来了：我们现在掌握了通往无穷的钥匙！剩下的就是使用这把钥匙来揭开关于无穷的秘密了。

下面是具体的做法。

我们已经定义了所有自然数的袋子，接下来，我们继续重复这个制作更大袋子的步骤，也就是把所有之前制作的袋子都装进一个新的巨大的袋子里。这个超级大的袋子里有多少小袋子呢？因为每一个自然数都是一个小袋子，所以这里面有无穷多个小袋子。

我们终于制作了无穷所对应的官方袋子：一个包含所有自然数袋子的大袋子。

换句话说，我们为无穷制作了一个官方袋子，里面装着无穷个小袋子，每个小袋子对应一个自然数。我们很快就会发现有一些无穷比这个无穷更大。这又是什么意思呢？

我们先练习几遍“一直数到无穷”的过程。记住，这不是说我们要大声地数出每一个自然数（因为我们的生命没有那么长），而是要把事物和我们的官方的无穷袋子里面的事物对应起来，对应方法就是双射函数。其实，我们在解决希尔伯特旅馆问题时已经这么做过几次了。把事物与无穷的官方袋子里的事物对应起来和把两层的希尔伯特旅馆里的客人重新安排到常规的一层的希尔伯特旅馆是一样的问题，关键在于：我们不能把两个客人安排进同一个房间，同时我们也不想浪费任何房间。

在第2章里，我们做过多次安排。我们已经明白了怎样把两层的希尔伯特旅馆、三层的希尔伯特旅馆甚至无穷层的希尔伯特旅馆的客人安排到一层的希尔伯特旅馆中。在这里，我们不再演示整个安排过程，只是简单地把旅馆中的客人与无穷的官方袋子里的物体相对应，看看是否对应得上。一个无穷的数集如果可以与无穷的官方袋子里的物体对应起来的话，那么我们就说这个无穷数集是“可数的”。但是我们已经有的那个无穷的官方袋子里的物体只包含自然数。所以，上面这句话也可以这么说：如果一个无穷数集能够和自然数对应起来，那么这个无穷数集就是可数的。或者说得正式一点儿，如果存在一个自然数集合的双射函数，那么根据这个双射函数得到的数集就是可数的。这类无穷数集叫作“可数无穷”——是的，我们在后面还会看到一些无穷数集属于“不可数无穷”。

现在，我们已经初步认识到有不同类型的无穷存在，因此我们要开始更加小心地书写无穷了。符号∞只是笼统地代表了非有限的事物。我们现在有了一个非常具体的关于无穷的概念——那个包含着所有自然数的无穷的官方袋子。数学家们有的时候把这个概念称为ω，也就是小写的希腊字母Ω。虽然Ω是希腊字母表中的最后一个，但ω恰恰是我们将要讨论的越来越大的无穷的一个开端。

令人惊奇的可数数集

一方面，你可能会觉得不可能有什么东西比无穷还大。另一方面，如果我们把另外一些东西扔进那个巨大的装着所有自然数的无穷袋子的话，那么它是不是还会变得更大？这个时候，我们就需要重新审视一下我们对于“更大”和“更小”的直觉了。因为当涉及无穷的时候，事情可能会和我们的直觉不太一致。如果我们试图去找一个不能和我们的ω号袋子对应起来的更大的无穷数集，我们会发现这相当困难。我们可能会在进行了各种各样的让这个袋子变得更大的尝试之后，发现它仍是可数的。

第一步，让我们检查一下这个包含所有自然数的数集，试着添加一些东西让它变得“更大”。比如，添加一只大象。我们想要看看这样一来这个数集是否依然可数：我们是不是能把这个数集里的物体和自然数完美地对照起来。

我们首先可以尝试把数集中的所有自然数与它们自身对照起来。等一等，大象哪去了？这看起来好像行不通。换一种方式，我们把大象对应到数字1，然后把所有的数字和比自己大1的数字对应起来。这样，1就对应到了2，2就对应到了3，以此类推。这样，所有的物体，包括大象，就都和自然数对应起来了，而且官方袋子里所有的自然数也都被用上了。

这个问题和客满的希尔伯特旅馆来了一个新客人的问题是一样的。在解决希尔伯特旅馆的问题时，所有其他的客人都往后挪了一个房间，如此一来，所有客人就都被安排妥当了。这个例子的意思是，如果你往一个无穷数集里放进一个额外的物体，实际上，这个数集并不会变得更大。这是我们关于无穷的一个基本直觉。换言之，“无穷加上1还是无穷”。现在，我们找到一些关于无穷在数学上的合理性了。

两层楼的希尔伯特旅馆就是两个自然数集。我们已经证明过包含两个自然数集的数集能够和一个自然数集形成双射函数（见图5–14）。这意味着，如果你把两个可数的无穷数集混合在一起，你得到的仍旧是可数的无穷数集。换言之，你还是没有得到“更多”的东西。事实上，我们也能据此推导出这个结论：无穷乘以2还是无穷。

现在，我们不用再考虑希尔伯特旅馆的问题了，我们可以开始考虑各种类型的数字，比如整数。整数几乎有自然数的两倍大，因为整数既包含正数也包含负数。那么整数是否理所当然地要比自然数多？事实上，如果一个希尔伯特旅馆既有正数房间也有负数房间，我们还是可以把这些客人安排到一个常规的一层的希尔伯特旅馆里。因为这并不会比转移两层的希尔伯特旅馆里的客人更加困难，而我们已经处理过两层的希尔伯特旅馆的问题了。整数的情况就相当于把所有的二层房间的号码全部标记为负数，更现

实地说，就像一个既有左翼又有右翼的旅馆一样。

那么有理数呢？有理数的数量是不是一定比自然数多？毕竟有理数在实数轴上的密度要比整数大得多。但事实上，如果希尔伯特旅馆的房间是用有理数编号的，我们还是能够将客人进行同样的安排。虽然这个过程有些复杂，而且包括不止一个步骤。

首先，记住每一个有理数都是能够写成$\frac{a}{b}$的形式的，其中a和b都是整数且b不为0。现在，我们先处理正的有理数。我们先把所有这些客人转移到无穷层高的希尔伯特旅馆的摩天大楼里，方法就是把$\frac{a}{b}$房间的客人转移到摩天大楼的b层a房间。这个办法是不是很巧妙？在完成这一步之后，我们就能按照老办法把他们都安排到仅有一层的希尔伯特旅馆里了。负数的有理数房间怎么办？很好办，我们可以把他们安排到另外一栋摩天大楼里，然后再转移到另外一个一层的希尔伯特旅馆。然后，我们就回到了两层的希尔伯特旅馆这个老问题上，而我们都已经知道该怎么做了。这是很典型的数学处理方式：将新的问题转换为我们已经知道该怎么解决的老问题。

你可能已经注意到了，我们将客人从有理数旅馆转移到摩天大楼旅馆的时候将会产生一些空房间。比如，因为$\frac{1}{2}$和$\frac{2}{4}$实际上是同一个数，所以，一旦我们把$\frac{1}{2}$号房间的客人转移到2层1号房间以后，就没有客人可以入住4层2号房间了。这说明我们所做的对应不是双射函数。但这并不是什么大事儿，只要我们所做的对应是内射函数就可以了。换言之，只要你还能把这些客人再安排回他们原来的房间那就不是什么问题。因为即便有一些房间空出来了，你总归还是可以通过让所有客人都往后挪一个房间来填充这些空房间。

所以，整数和有理数并不比自然数多，虽然它们看上去要多得多。但要注意一点：这并不意味着所有的无穷数集都是一样大的。你可能已经注意到了，我们还没有检查实数有多少个。如果我们把无理数加到有理数的袋子里，会发生什么呢？答案是，事情终于发生了变化，现在我们确实有更多的物体了。在接下来的一章里，我会用这个事实来证明一个看起来似是而非的论断：一些事物比另一些事物“更无穷”。

如果你看到湍急的河流中出现了一个泛着白色水花的漩涡，你是不是会产生一种冲动——想要坐上一个筏子然后猛地冲进去？如果我面对的是一条

真正的河流的话，我大概不会这样想。但如果我面对的是一条数学之河的话，我就一定会这么做。我不太喜欢在现实中被搞得晕头转向，但是我喜欢我的大脑为了弄明白那些最初看起来几乎不可能理解的事物而疯狂运转的状态。验证“一些事物比另外一些事物更无穷”这个事实就是这样一个激动人心的数学问题。做好准备了吗，我们要跳进这个漩涡了。

6 接近无穷

孩子们有时会发生这样的争论：

“我是对的。”

“我比你更对。”

“我比你对一百倍。”

“我比你对一万倍。”

“我比你对一亿倍。”

“我比你对无穷倍！”

“我比你对两倍无穷。”

“我比你对无穷的平方！”

然而，我们似乎已经意识到，无穷的两倍并不比无穷更大，甚至无穷的平方，也就是无穷乘以无穷，也不比无穷更大。所以最后，哪个孩子也没能比另一个更正确。在这场争论中，当一个孩子说出他比别人对无穷倍以后，其他的孩子们还能否做点儿什么打败他呢？答案是：能。他们只需要找出一个比那个孩子的无穷更加无穷的无穷！我们将采用两种方式构建比自然数“更加无穷”的事物。第一种方式是试着数无理数。

无理数比有理数更多

在这个世界上，不理性的人似乎比理性的人更多。事实上，大多数人都多少有点儿不理性。在我看来，这是人和计算机的一个重要区别。在人的一生中，完全理性并始终遵循逻辑是非常困难的。感情是非理性的，口味也是非理性的。你肯定有自己更喜欢吃的食物。有的时候，这种偏好是可以用理性来解释的。例如，我不喜欢很辣的辣椒，因为它会让我的嘴感到很疼。但是我为什么不喜欢肉桂呢？我也不知道，我就是不喜欢它。当人们发现有人毫无理由地喜欢或讨厌什么的时候，他们通常就会将其视作一种完全非理性的行为。当我发现有人不喜欢巧克力的时候，我的第一反应就是这样的。然后，我会提醒自己，这只是一种口味问题，而且我认识的不喜欢巧克力的人要比像我一样不喜欢肉桂的人还多。无论如何，即使你不认为自己是非理性的，你也肯定不是纯粹理性的。我们的日常用语比数学

用语存在更多的灰色地带。事实上，学习数学的目的之一就是消除灰色地带——当然，不是消除整个世界的灰色地带，因为这是不可能的，也并不会令人感到愉悦。不过我们可以从我们现在所想的做起。

世界上的无理数比有理数更多。事实上，让一个数字成为有理数是一件很困难的事情。成为两个整数的比值是一个非凡的巧合。如果你觉得这很奇怪，可能是因为我区分了“数学上的可能”和“人类意识上的可能”。

如果有人在街上拦住你并让你想一个数字，你很可能会想到一个有理数。事实上，我猜你很可能会想到一个正整数（虽然也有少部分人会提到π）。但这并不意味着正整数比其他类型的数更多，只是意味着我们的大脑更熟悉正整数。毕竟，正整数被称为自然数。

然而，如果我现在要求你想一个数字，并以它为半径画一个圆，那么这个圆的面积的数值几乎可以确定是一个无理数。还记得吗，半径为r的圆的面积是$\pi r2$。因为π是一个无理数，所以如果这个圆的半径是一个整数或者任何一个有理数，那么其面积一定是一个无理数。

如果你用一个有理数乘以一个无理数，其结果将是无理数。同样的，如果你用一个有理数加一个无理数，其结果也将是无理数。但是如果你把两个无理数相乘，则其结果可能是有理数也可能是无理数。例如，$\sqrt{2}\times\pi$的结果是无理数，但$\sqrt{2}\times\sqrt{2}$的结果就是有理数。

你能不能找到一个值使得以其为半径的圆的面积为有理数？我们应该怎样消除公式里面令人讨厌的π呢？下面我们将会尝试消除它。如果我们把半径的值设成$\frac{1}{\sqrt{\pi}}$，那么圆的面积就是：

$$\pi \times \left(\frac{1}{\sqrt{\pi}}\right)^2 = 1$$

我猜，如果我在街上拦住你让你想一个数字，你说的数字应该不会是$\frac{1}{\sqrt{\pi}}$。

当然，经过现在这些推算，在将来被问到的时候，你可能就会想到说这个数字了。

还有其他的值能使以其为半径的圆形的面积是有理数吗？让我们回忆一下，有理数是能够表达为$\frac{a}{b}$的数，其中a和b都是整数且b不等于0。所以，让我们尝试把$\frac{a}{b}$变成圆的面积。我们可以将半径设置为：

$$r = \sqrt{\frac{a}{\pi b}}$$

这样我们就能得到圆形的面积：

$$\pi r^2 = \pi \times \left(\sqrt{\frac{a}{\pi b}}\right)^2 = \frac{\pi a}{\pi b} = \frac{a}{b}$$

综上所述，我们会发现无理数比有理数多得多，只不过我们更容易想到有理数。但是，这个结论似乎并不严谨。我含糊地扔出了“很可能”、“几乎可以确定”和“多得多”这样的表达方式。通常在数学领域里，我们被允许在最开始的时候使用这种模糊的语言描述我们的直观感受，而经过一段时间的学习后，我们就要将其雕琢成符合逻辑的数学观点。我们接下来要做的就是雕琢。

如此多的无理数

虽然我们还无法准确说出无理数是什么，但不管怎样，接下来，我们要开始“数”无理数了。无理数的定义是通过说明它们不是什么来给出的：它们是不是有理数的实数。我们能够准确说出实数是什么（尽管在第15章之前，我们对此只能有一个模糊的感觉）。我们也能够准确说出有理数是什么。所以用排除法来“数”无理数比直接“数”无理数要容易一些。简单来说，我们要做的就是，“啊，实数好多啊！但是其中的有理数好像不是很多。所以如果排除其中的有理数，那么剩下的就是无理数了。”我们只需要比这做得更精确或更具技术性一点儿。

在证明无理数是不可数的之前，我们需要先来证明实数是不可数的。我们已经知道以下事实：

* 有理数是可数的。

* 如果你把两个可数的数集合并在一起，你就会得到另一个可数的数集（就像是两层楼的希尔伯特旅馆一样）。

这告诉我们，如果无理数是可数的，那么实数也是可数的。而接下来，我们会证明实数是不可数的，因此无理数也不可能是可数的。

这有点儿类似于显性基因和隐性基因的表达关系。想象一下，我们把不可数性比作显性基因，把可数性比作隐性基因。如果你把可数的数集和不可数的数集放在一起，你会得到一个不可数的数集，因为不可数性是显性基因。如果你把一个可数的数集和一个未知的数集放在一起，并且你知道最终得到的数集是可数的还是不可数的，那么你就能推断出未知的数集是可数的还是不可数的：

有理数	如果实数	那么无理数
可数（已知条件）	可数	可数
	不可数	不可数

实数是不可数的

实数是很难界定的。但是我们暂时可以认定它为一切能表达成小数的数的总称。小数的位数可以是无穷的，无论是否循环。事实上，如果可以在后面加0的话，所有的小数都将永远没有尽头，我们只是通常不写这些0而已。我们已经能很容易地想象出将小数位数展开到无穷的样子，所以我们现在大概很难理解为什么不能直接参照它来定义实数。这个问题我们会在本书后文给出解答。在第14章我们将会讨论无穷小的事物并运用相关知识深入探索这个问题。在此之前，我们先来看看我们能解决的问题。

我们将要证明这些有无穷位数的小数是不可数的，使用的方法就是格奥尔格·康托尔提出的一个聪明的技巧。这个方法最广为人知的名字是康托尔对角线论证。事实上，我们将要证明0到1这个区间内的实数就已经是不可数的了（将注意力集中在0到1这个区间内的数字上能够简化整个证明过程）。

让我们换一种方式来思考这个问题。我们将再次思考怎样转移巨型的无穷旅馆的客人这个问题。在第2章，我们提到过实数旅馆的想法。在这个旅馆里，每个实数对应一个房间。现在我们稍微谦虚一点儿：假设有这样一个没那么夸张的旅馆，这个旅馆“只有”与0到1这个区间内所有实数相对应的房间。虽然我用了“只有”这个词，但这仍然是一个巨型的无穷旅馆。这将是我们遇到的第一个不能将客人安排到常规的一层的希尔伯特旅馆的巨型无穷旅馆。

只考虑0和1之间的数字是一个很便捷的方法，因为我们不用担心小数点前面的整数部分把问题复杂化。每个房间都对应一个不同的小数，这个小数的展开式以“0.”开始且没有终点。我们现在忽略你需要花多少时间才能找到你的房间钥匙这个问题，先来想一下怎样转移这家旅馆里的客人。换言之，这个问题的实质是，怎样将0到1这个区间内的实数和自然数完美配对——这正是我们所说的可数的定义。

我们可以从尝试这几种方法开始。

我们是不是可以按照房间号的大小顺序让客人住进新的房间呢？即让原本住在最小的小数号房间的客人入住1号房间，然后以此类推。不对，这不行。因为就像没有最大的数字一样，并不存在最小的小数。0.0000000000001是最小的小数吗？不是，因为我们总是可以在这个数中间插入更多的0让它变得更小。

我们是不是可以从小数点后只有一位的小数开始处理，然后再处理小数点后有两位的小数，以此类推？毕竟，每一位上的数字都是有限的。比如，小数点后只有一位的小数只有10个，就是0.0、0.1、0.2、0.3……0.9。小数点后有两位的小数呢？十分位有10种选择，百分位也有10种选择，所以这样的小数一共有$10\times10=100$个，如下所示：

0.01	0.11	0.21	0.31	0.41	0.51	0.61	0.71	0.81	0.91
0.02	0.12	0.22	0.32	0.42	0.52	0.62	0.72	0.82	0.92
0.03	0.13	0.23	0.33	0.43	0.53	0.63	0.73	0.83	0.93
0.04	0.14	0.24	0.34	0.44	0.54	0.64	0.74	0.84	0.94
0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
0.06	0.16	0.26	0.36	0.46	0.56	0.66	0.76	0.86	0.96
0.07	0.17	0.27	0.37	0.47	0.57	0.67	0.77	0.87	0.97
0.08	0.18	0.28	0.38	0.48	0.58	0.68	0.78	0.88	0.98
0.09	0.19	0.29	0.39	0.49	0.59	0.69	0.79	0.89	0.99

小数点后有三位的小数更多，但仍然是有限的。这个方法看起来可行，但还是存在一个问题：我们只能转移那些房间号对应的小数有有限位数的客人，而永远也转移不完那些房间号对应的小数有无穷多位数的客人，哪怕我们以这个方式“永远”地做下去。这有点儿像，如果你给一个数字不停地加1，那么你得到的数字会越来越大，但是结果永远也不会真的变成无穷。同样的，如果我们继续用这种方式不停地转移客人，我们处理的小数就会越来越长，但是给小数展开式的位数不停地加1，其结果同样永远不会真的变成无穷。

如果你不相信这个结论，那也没什么，因为这并不是一个数学论证，所以你没有被说服是正常的。但是康托尔对角线论证真的是一个无懈可击的数学论证。我们必须证明，无论我们付出多大的努力，都不可能将这个巨型无穷旅馆里的每一位客人都转移到一层的希尔伯特旅馆里。这个论证听起来好像特别困难，因为我们好像必须证明每一种转移方法都会失败。康托尔的方法的聪明之处在于，你根本不用进行任何实际的转移工作。你只需要假设我们已经完成了这个任务，也就是说，假设现在所有客人都已经被转移到一层楼的旅馆里了，那么你接下来要做的就是找到其中的矛盾之处。我们需要证明的是，如果任何自作聪明的人宣称他已经把所有的客人都安排好了，那么我们总能发现至少还有一位客人被落下了。从数学的角度讲，这相当于说如果我们声称我们把所有的小数都和自然数匹配上了，那么一定会有至少一个小数没有与之相匹配的自然数。

下面是这个方法的工作原理。第一步，你要到1号房间敲门问这位客人他原本的小数房间号是多少。事实上，你并不需要问出这个小数的全部展开式，你只需要问第一位。（这点儿像你登录某些网站的时候，网站只要求你写出密码的第3位和第7位字母或数字。）在他告诉你他原本的房间号的第一位小数之后，你把这个数字加1然后写下来。比如，如果他原本的房间号的第一位小数是3，你就写下4。如果这个数字是8，你就写下9。但如果数字是9，你不能写下10，因为10不是一位数，这时你要写下0。这个方法现在听起来还很令人费解，但请继续往下听。

第二步，你要去2号房间问这位客人他原本的小数房间号的第二位数字。和前面一样，在这个数字上加1然后写下来，注意紧挨着你前面写的数字。

第三步，你要去3号房间问这位客人他原本的小数房间号的第三位数字。和前面一样，在这个数字上加1然后写下来，同样紧挨着你前面写的数字。

我们正在做的事情是构建一个新的小数展开式。你需要问住在新的旅馆n号房间的客人其原本的小数房间号的第n位数字，然后加上1得到这个新的小数展开式的第n位小数。

如果你在“现实生活”中采用了这样的做法，你将永远完成不了这项工作。但是数学论证并不要求你真正去敲每个房间的门。无论你是不是真的花了时间写下它，我们在这里建立的数字都是存在的。就像在人类认识海王星之前，海王星就已经存在了一样。

问题出现了：曾经住在我们通过加1构建的这个小数房间的人去哪儿了？这位客人被转移到了哪个编号为自然数的房间？

这个人不可能在1号房间，因为这个新的小数和1号房间的客人原本的小数房间号的第一位数字不符。这个人也不可能在2号房间，因为这个新的小数和2号房间的客人原本的小数房间号的第二位数字不符。这个人也不可能在3号、4号、5号或者任意n号房间，因为这个新的小数和n号房间的客人原本的小数房间号的第n位数字不符。如此我们便找到了一位客人，这位客人没有被转移到这个一层的希尔伯特旅馆的任何房间。这样，我们就找到了这个假设的一个矛盾，证明了那个自作聪明的人是错的：他并没有成功转移每一位客人。无论他采用了什么样的转移方式，这个办法都能够证明转移是不成功的。

这种方法被称为对角线论证的原因是，如果我们把这些小数房间号按照位数对齐写成一列，我们要关注的就是它的对角线。让我们假设前几个房间

看起来像下面这样：

新房间号	旧房间号
1	0.238795317…
2	0.984718573…
3	0.389716438…
4	0.777362889…
5	0.444317895…
6	0.879000001…
7	0.892225673…
8	0.191919234…

当我们敲每个房间的门时，我们需要的数字就是加下划线的数字。可以看到，这些数字都在一条对角线上。

新房间号	旧房间号
1	0.238795317…
2	0.984718573…
3	0.389716438…
4	0.777362889…
5	0.444317895…
6	0.879000001…
7	0.892225673…
8	0.191919234…

如此，我们构造出来的新数字的前几位小数就是下面这样的：

0.39042174…

我们可以通过观察它的第n位数来证明这个人不在n号房间里，因为这个数字和住在n号房间里的客人其原本的小数房间号的第n位不同。所以，原本住在上面这个小数房间号的客人没有被转移到一层旅馆的任何房间，那个自作聪明的人失败了。

我们用这个方法证明了0到1这个区间内的实数是不可数的。如果你试图把这些实数和自然数完美配对的话，那么至少有一个数字会被落下。据此，我们可以认为实数比自然数“更无穷”。

这种证明方法打败了那些认为自己可以投机取巧的自作聪明的人。小心，

这些自作聪明的人在我们研究无穷小的时候还会再来找我们的麻烦。

决策疲劳

还有另一种形式可以用来表达某些事物比自然数“更无穷”。这种形式可能和“不可数”这个词联系得更为紧密。我们不能把实数旅馆里的客人转移到自然数旅馆里，因为总是有人会被留在寒风中。然而另一种证明我们无法转移这个旅馆的客人的方式叫作决策疲劳。想象一下，你有一个无穷旅馆，里面全是双人房间，你需要把客人都转移到一个全是单人房间的旅馆里。你可以将原本住在1号房间的两位客人转移到1号房间和2号房间，将原本住在2号房间的两位客人转移到3号房间和4号房间。以此类推，原本住在n号房间的两位客人会被转移到2n号房间和2n–1号房间。表面上看，这和把客人从两层楼的希尔伯特旅馆转移到一层楼的希尔伯特旅馆是一样的，但是这次你要怎么写转移指南呢？

“如果你住在n号房间，那么你们中的一个人去2n号房间，另外一个人去2n–1号房间。”

如果他们问“我们哪个人去哪个房间呢”，你可以说“年纪大的去2n号房间，年纪小的去2n–1号房间”。但是如果偶数号房间都比奇数号房间好的话，年纪较小的客人是不是会觉得很不高兴呢？或者，请允许我做一个不太可能的假设，如果所有房间里的两个人年纪都一样呢？你可以让他们自己决定，因为结果并不重要。但是万一他们优柔寡断，想要你来帮忙决定呢？显然，这个假设会在某个地方无法继续下去，而我的意思是，写下指南的意义在于让人们可以简单地遵循说明而无须思考。如果他们需要在转移过程中做一些决定的话，那么这个指南从数学上讲就是一个含糊不清的指南。换句话说，它不是一个计算机可以遵从的指令，或者说不是一个合规的算法。

这种情况也可以被比喻成数鞋子和数袜子的差别。假设你有无穷多双鞋，而且是可数的无穷。你可以把它们按顺序放在编号为1、2、3……的鞋盒里。那么，这是否意味着单只鞋也是可数的？如果你有无穷双袜子，那情况又是怎样的？

我不经常数我的鞋，因为我可能会被鞋子的数量震惊到。我的借口是我的脚很大，很难找到合适而且不丑的鞋。我以前很胖，而那个时候，我的脚比现在还大。所以我的习惯是每当遇到合适的、不丑并且不贵的鞋，我都会买下来。那个时候，符合所有这些条件的鞋并不是很多，所以全部都买下来是可行的。可以这么说，我对于买鞋有一套自己的标准。事实上，直到毕业，我仍然只有大约四双鞋（虽然这已经比我认识的一些男性的鞋子要多了）。在我的体重减轻后，我发现我的脚也变小了（我之前并没有意

识到脚也会发胖）。虽然只小了半个鞋码，但是这关键的半码让我能买的鞋从超大码变成了正常码。但是，我买鞋的习惯并没有变，这意味着我最终买了过多的鞋子。我觉得这比冲动购买跑车还要稍好一些。当然，也有可能是我说得太过了。总之，现在我必须在买鞋子的时候做出有针对性的决定，而不是简单地根据习惯购买。

然而，无论我买了多少双鞋，我的鞋子的数量都是有限的。但是让我们想象一下无穷双鞋，这些鞋子整齐地摆放在按自然数编号的盒子里（当然，我的鞋子从来没有整齐摆放过）。

这意味着我们有可数“双”的鞋子。但是我们的鞋子的“只”数是可数的吗？也就是说，我们是否能够把无穷只鞋整齐排列并让每只鞋匹配一个自然数，而不是让每双鞋匹配一个自然数？是的，我们可以做到。这和转移两层楼的希尔伯特旅馆里的客人是一样的。我们可以决定每次都从左脚开始，排列起来就是：1号左鞋、1号右鞋、2号左鞋、2号右鞋、3号左鞋、3号右鞋……以此类推（见图6–1）。写成符合数学格式的说明，就是：

- n号左鞋放在位置2n–1上。

- n号右鞋放在位置2n上。

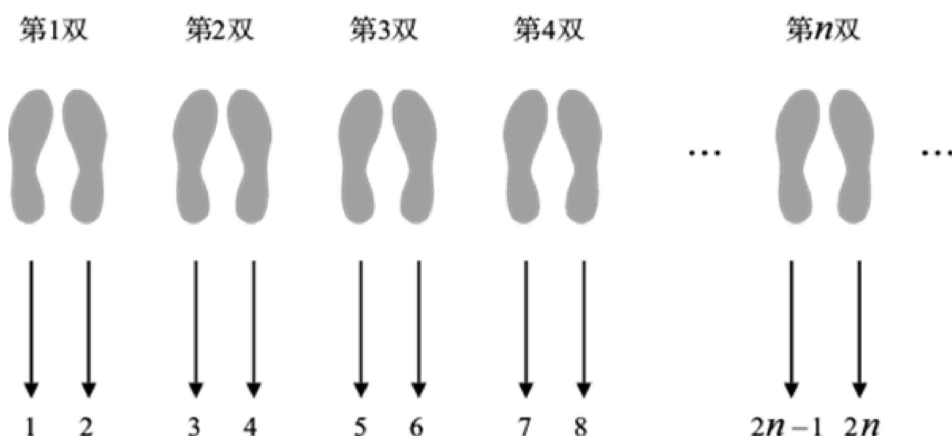


图6–1

这就像是我们之前做过的将由红色自然数和蓝色自然数组成的集合对应到紫色自然数上一样。在这里，“左鞋”相当于蓝色自然数，“右鞋”相当于红色自然数。

现在，让我们思考一下袜子的情况。假设我们有无穷双袜子。我们把它们和鞋子一样排成一排：第1双袜子、第2双袜子……以此类推。（我对袜子的兴趣不像我对鞋子的兴趣那样大。我所有的袜子都是黑色同款的，所以我不用考虑将它们配对。）回到这个问题上，我们是不是能够将袜子按照只数排成一排呢？乍看之下，这个问题和鞋子的问题是一样的。但是我们遇到了一个问题——并不存在左袜和右袜这样的说法，因为左脚的袜子和右脚的袜子看起来是一样的。我们是不是能够指定哪只排在前面呢？答案是，我们不能。虽然我们可以武断地把一双袜子排在另外一双袜子前面，但是我们必须依次针对每双袜子做决策。对于鞋子，我确实也做过武断的决定：左鞋排在前面，右鞋排在后面。但是我只需要做这一次武断的决定就可以了。而对于袜子，每当有一双新袜子，我都必须做一次新的决定，指定这双袜子的哪只排在前面，哪只排在后面。此时就产生了决策疲劳的问题，至少是数学上的决策疲劳。

决策疲劳的理论就是说，无论是小的决定还是大的决定都会让人觉得疲劳。前者如，我应该吃什么早餐？后者如，我应该买哪一幢房子？在一天之内做的决定越多就会越让人感到疲劳。做决定是很难的。你可以衡量利弊，但是最终你必须迈出做决定的那关键一步。仅靠遵循逻辑不能帮助你得出结论，因为那样的话你得出的只是一个推断，而不是决定。

就像数袜子一样，我们也有一个数学版本的决策疲劳。我们可以做出一个武断的决定，我们也可以做出两个、三个或者任意有限个武断的决定。但是我们能否做出无穷多个武断的决定呢？在袜子的例子中，这个问题就等同于我们是否能写下一个公式将袜子排成一排。从技术上讲，我们必须写出一个函数将袜子集合对应到自然数集合。这有点儿像是我们要把每一只袜子都疏散到一层的希尔伯特旅馆里一样。但是我们怎样表述这个函数呢？如果一双袜子中的两只之间没有任何的不同点，我们怎样描述一双袜子中的哪一只应该排在前面呢？

你可能会想到用下面这种办法来做决定：把袜子拿在手里，一手一只，左手的那只排在前面，右手的那只排在后面。这种做法好像有点儿道理，但在数学上仍然不能算是一个足够好的说明。“决定”并不是那个正确的词，“选择”才是。即使你随机地拿起袜子，在某种程度上，你还是面临着用哪一只手去拿哪一只袜子的选择。换言之，针对无穷双袜子，你还是得做出无穷次的选择。

做出无穷次选择是否可行这个问题在数学上还没有得到明确的解决，这个微妙的问题仍在困扰着数学家。当前，它被称为选择公理（Axiom of Choice）。公理是一个你认定它是正确的且无须证明的基本假设，也是构建所有其他事物的基石。关于公理的一个思考方法是，你不能只是简简单单地说它是正确的，你还要以它为基础构建一个世界，然后研究看看这个

世界中会发生什么。

选择公理是指，做出无穷数量的武断选择是可能的。在选择公理是正确的世界里，袜子是可数的。但是在选择公理不正确的世界里，袜子是不可数的。我们也可以利用同样的原理处理有无穷多个双人房间的旅馆问题。如果一个房间里的两位客人不总是同样的年纪的话，我们就可以利用一位客人年纪大一点儿而另外一位客人年纪小一点儿的这个事实去安排房间，从而避免每次都要面对哪一位客人应该转移到哪个房间的选择。

数学家们完全不在乎选择公理是否总是正确的，他们只关心我们在给定的情况下是不是需要使用这个公理。如果我们在每次使用到这个公理的时候都要指出来这个公理可能存在的问题，这就会让人觉得非常麻烦。这有点儿像我们每次倒车时警报系统发出的嘟嘟声。

因此，这给了我们另一种不可数的形式：不是因为数量太多，而是因为事物之间不可区分。当我在一个房间里面数人的时候，偶尔也会出现这种情况，比如我在教室里和一群还不认识的新同学在一起的时候。我希望你也有过这样的经历，那么这就不是我不会数数的问题了。如果每个人一排排地整齐坐好，那就没有问题，因为你可以一排一排地数。如果大家都随意地坐在摆放得乱七八糟的椅子上，那么数起来就困难了，因为你必须决定下一个数谁。通常，我发现数9到10个人是没有问题的，如果再往上数的话，我就会把已经数过的人和接下来要数的人搞混。想象一下要在后一种情况下数到无穷：即便不考虑我们在有限的人生里是否有足够的时间做这件事，数起来也是极为困难的。

的确，我也听过很多说数学家不会数数的笑话。但事实上，数数是一门高深的学问，这里的数数和我们说一个孩子“会数数了”的时候所说的那个“数数”有着本质区别。在聊天的同时做咖啡，我就会数不清放了多少勺咖啡粉。从这个问题到不可数的无穷数集的问题，这些问题对于数学家来说都是拓展思路的肥沃土壤。现在我们已经知道，实数比自然数多，在下一章，我们将会试着数一数实数，看看它们到底有多少。同时，我们通过本章的练习得到提升的洞察力将带领我们走进层级越来越多的无穷世界。最终，我们会为那些试图证明自己比小伙伴的“对无穷倍”更对的孩子找到答案。

7　超越无穷

我最近参观了美国新墨西哥州圣塔菲附近的帐篷岩国家纪念地。这次徒步旅行实在令人激动，我穿过狭窄峡谷中的锥形岩层，最终爬上了山脊，看到了里奥格兰德河谷壮观的景色。这是一次令人满意的徒步旅行，一路上，我近距离观赏了帐篷岩石，而最终到达山脊时看到的景观简直有一种戏剧性的效果，地平线以上的景物一览无遗。一路上，我曾经数次感到恐惧。我并不是一个特别勇敢的人，站在岩石上的时候，往下看对我来说有点儿太恐怖了。特别是当我们走在特别陡峭而岩石又非常松散的路段时，我总会觉得自己没有找到好的立足点。当我到达山顶时，我享受到了短暂的成就感。而在这之后，我立马意识到还有更多的路在等着我。它不是一个单独的山峰，而是一个山脊，所以我们可以沿着它一直走。看着两边的悬崖峭壁，我考虑了一下要不要继续，然后感到走到这里的我已经足够有勇气了，便决定不再继续走下去了。

当我们攀登无穷这座高峰的时候，你可能也会有类似的感觉。一路上，你可能都觉得自己没有什么稳固的立足点，而现在，我们终于到达了无穷的一个山头，但是你可能已经想要离开了。我在面对数学时总是比在面对山峰时更加勇敢，所以让我们继续走下去。我们已经看到，实数的数量比自然数的数量更加无穷，但是实数的尽头在哪里呢？从一个无穷走向另外一个更大的无穷到底需要多久？下面，我们就来数一数。

当然，我们不可能把实数一一列举出来，然后数一数一共有多少个。相反，我们会像前面几章数其他事物一样数实数。换言之，我们将抽象地数。我们会想一想实数是怎么构建出来的，然后找到它们和自然数之间的关系。这样，我们就能弄清楚两者之间相差多远。首先，我们会数一些小一点儿的数集来让自己习惯这种抽象的数数方式。然后，我们就可以用这种方式来数无穷了。

抽象的数数

看一看下面这个套餐菜单：

每日例汤

∞

烤三文鱼配柠檬牛油汁

或

香草烤鸡配土豆泥

∞

巧克力蛋糕

或

柠檬挞

上面的菜单中包含多少种可能的晚餐形式？我们可以通过列清单的方式将每一种组合具体地写下来：

* 三文鱼和蛋糕

* 三文鱼和柠檬挞

* 鸡肉和蛋糕

* 鸡肉和柠檬挞

因为我们给出的菜单很短，所以目前列举起来还是很容易的。现在，想象一下一个十道式的套餐，每一道菜都有三个选择。要把在这种情况下所有可能的组合的数量算清楚，就需要花上很长一段时间了。这个时候，抽象的数数方法就派上用场了：我们可以靠推理数数，而不是靠依次列举出第1、2、3、4……种方案。作为过渡，我们可以先将上述菜单中的所有可能性画成下面的树状图（见图7–1）。

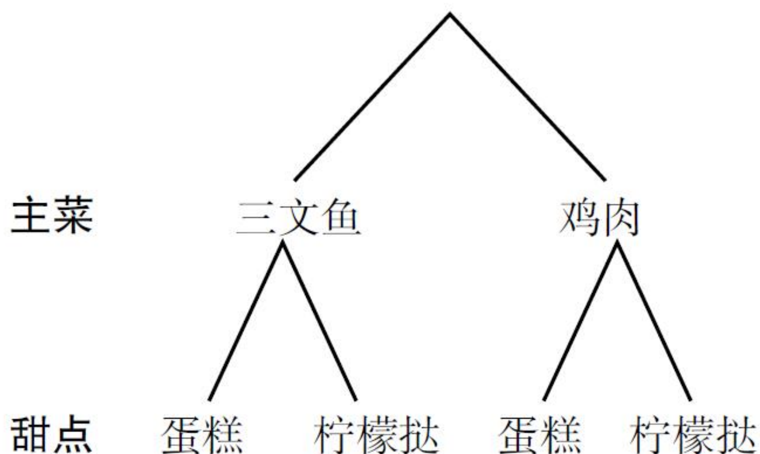


图7–1

树状图的顶层表示我们有两个可能的主菜选项，底层表示无论选择哪个主菜，我们都有两个甜点的选项。从上到下的每一条路径都表示一个可能的菜单组合，而这个树状图一共有4条路径。我们也可以抽象地说，顶层的两个选择通过底层的两个选择得到倍增。通过这种方法，我们可以看出最后的组合数是2×2＝4种。这个时候，我们可能会疑惑4是由两个2相乘得来的，还是由两个2相加得来的。因为2＋2也能得出同样的答案。但是，让我们给前菜加入另外一个选项再来看一看（见图7–2）。

每日例汤

或

蔬菜沙拉

∞

烤三文鱼配柠檬牛油汁

或

香草烤鸡配土豆泥

∞

巧克力蛋糕

或

柠檬挞

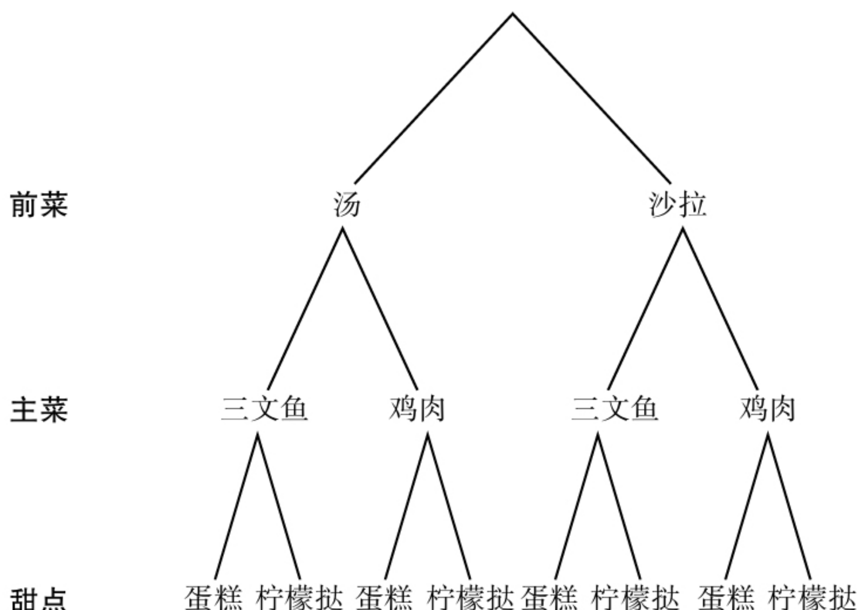


图7–2

从图7–2可以看出，这一次，顶层的两个选择通过第二层的两个选择得到倍增，两者的乘积再通过底层的两个选择得到倍增。所以可能的组合数是$2\times2\times2=8$。

我们也可以用类似的方法来数小数。在上一章里，我们通过研究0到1这个区间内的实数来研究所有的实数。这里，我们也用同样的办法先来数一数0到1这个区间内的所有实数。之前，当我们试图把那些房间号为小数的旅馆里的客人按照一位小数一位小数的方式转移的时候，我们没有成功。但是这里，我们再尝试一下同样的办法。我们先来看第一位小数，这有点儿像菜单中的第一道菜。只不过，我们面临的不再是2个选项，而是10个选项，因为可能出现在小数十分位上的数字有10个。只有一位小数的10个可

能的数字就是下面的这些：

0.0、0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9

一个有两位小数的数字就像一个两道式的菜单，其中每道菜都有10种选项，所以一共有$10\times10=100$种可能的组合。根据树状图分析可得，第一层有10个可能性分支且每一个分支都连着10个另外的分支，所以最后总共会有100片或是10^2片叶子。我们在前面一章已经看到过，0和1之间有两位小数的100个数字就是下面这些：

0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.19
0.20	0.21	0.22	0.23	0.24	0.25	0.26	0.27	0.28	0.29
0.30	0.31	0.32	0.33	0.34	0.35	0.36	0.37	0.38	0.39
0.40	0.41	0.42	0.43	0.44	0.45	0.46	0.47	0.48	0.49
0.50	0.51	0.52	0.53	0.54	0.55	0.56	0.57	0.58	0.59
0.60	0.61	0.62	0.63	0.64	0.65	0.66	0.67	0.68	0.69
0.70	0.71	0.72	0.73	0.74	0.75	0.76	0.77	0.78	0.79
0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99

我们现在把这个规律推广至有任意n位小数的数字，就像一个n道式的菜单或者有n层的树状图。对于有n位小数的数字，一共有$10\times10\times\ldots\times10$种可能性，也就是$10^n$种可能性。

现在，只要向前跳跃一小步，我们就能得出下面的推论：对于有无穷位小数的数字来说，它一共有“10的无穷次方”种可能性。虽然我说这是跳跃一小步，但实际上，这是集合论更深层次的内容。我们数实数，实际上是试图找到一个官方的袋子，这个袋子里面的物体能够和实数一一配对。你可能认为我们可以简单地说“这是实数集”，就像我们定义自然数集的时候一

样。但是如果我们能够把这个集合和自然数集合联系起来，那将会有更大的意义和启发性。“10的无穷次方”这个描述已经很接近了。但是为了能够将这个概念描述成一个事物的集合，也为了能够得到更加令人满意的答案，我们最好把这个问题转换为二进制。

关于二进制的一个稍稍离题的讨论

二进制是一种数字系统。在十进制里，我们通常使用0、1、2、3……9这些符号表示数字。但是在二进制里，我们只使用0和1。仅仅用0和1来封装的信息可以多到惊人。计算机基本上就是依靠这个系统来运行的，每个0和1都是一个开关，而计算机系统里有数万亿个这样的开关。你可以用为数不多的开关得到很多不同的配置。

用两个开关，你可以得到4种配置：

1 号开关	2 号开关
关	关
关	开
开	关
开	开

用三个开关，你可以得到8种配置：

1号开关	2号开关	3号开关
关	关	关
关	关	开
关	开	关
关	开	开
开	关	关
开	关	开
开	开	关
开	开	开

这就像我们之前讨论的两道式菜单和三道式菜单。如图7–3所示，这些开关构成的树状图看起来就像下面这样：

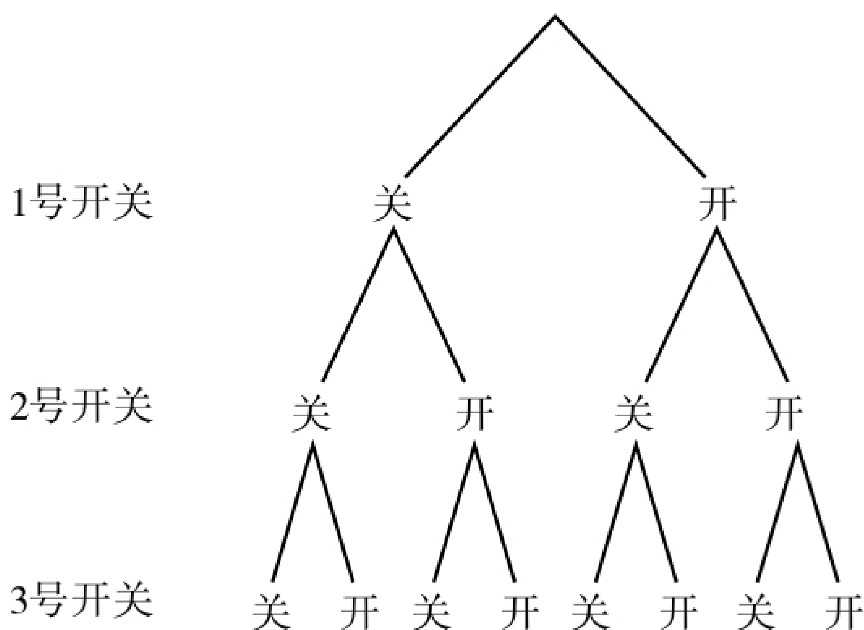


图7–3

树状图最底层的每一个节点都代表一种这三个开关能提供的配置形式。在这里，我们一共有8种可能的配置。我们只需要遵循从树状图顶部到底部给定节点的路径，就能记录下经过的每一个开关是“开”的状态还是“关”的状态。

虽然我们把这些树形的图示称为树状图，但是它们看起来有点儿奇怪，因为在正常生活中，树是从底部开始向上生长的。但是，数学上的树一般是从顶端开始往下生长的，大多数时候，这是因为我们的阅读习惯是从上到下的。当然，也有一些人会把树状图侧过来画（见图7–4）。

我希望你能明白，用哪一种方式画树状图并不重要。因为无论用哪一种方式表达，我们编码的信息都是相同的。我们可以把这种抽象而非具象地描述事物相互关系的图形称为示意图。随着数学越来越抽象，示意图的作用也越来越得到凸显，因为事物之间的抽象关系变得更加微妙和重要。而且，示意图对于情形的概括往往比言语描述更加简洁明了。一个具体的例子就是我们用来描述如何转移希尔伯特旅馆的客人的示意图。

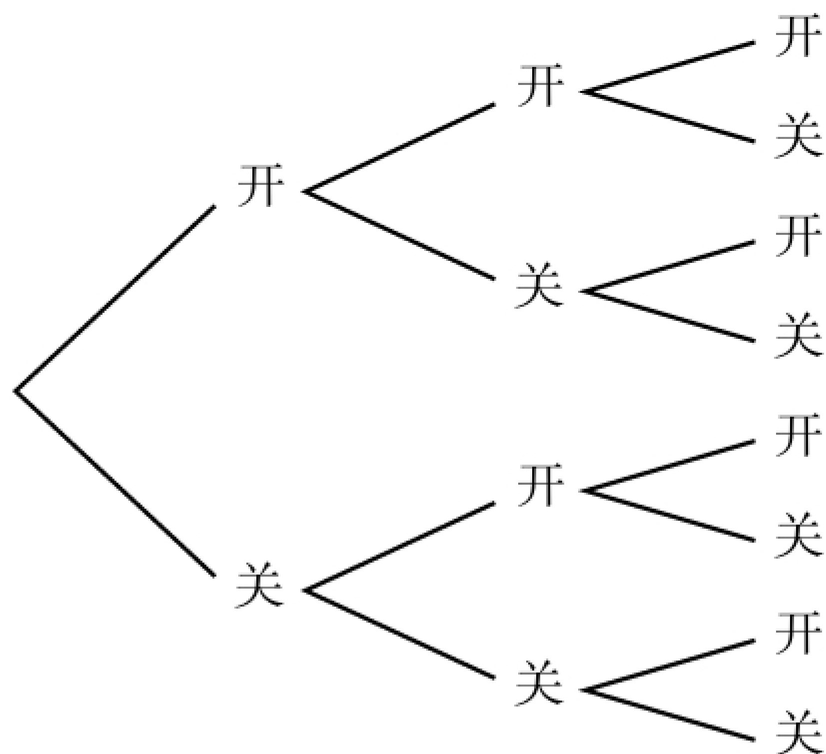


图7–4

基础数学大部分都是线性的。例如，求和的表达式是：

$$3+2=5$$

方程式是：

$$2x+3=7$$

这些符号都整整齐齐地待在一行里和睦相处。即便是在计算过程中，我们也会用一行一行的表达式来表述计算步骤：

$$2x = 7 - 3$$

$$2x = 4$$

$$x = 2$$

我们在后面会看到，随着数学的发展，数字的维度会逐步增多。如果我们研究的物体具有形状，那么它们结合的方式就会超出线的范畴。这有点儿像装配拼图块或者搭建复杂的乐高。想象一下，你正在用语言向另外一个人解释如何用乐高搭建一辆小汽车，而不是使用图形说明书来完成这项工作。就像一张图能够包含千言万语一样，一个数学示意图能够迅速而生动地描绘一个场景。

在上面的树状图中，我们可以看到，我们每增加一个新的开关，就要将树状图尾部的节点再分成两个枝杈构成新的一层。这些尾部的节点常常被称为“叶子”，因为虽然树状图的树是倒着生长的，但是这些节点仍然处于枝杈的最末端。

我们每添加一个新的开关，都相当于把可能的配置数量翻了一倍。意思是，如果我们有4个开关，那么我们可能的配置总数就是$2\times2\times2\times2$，也就是2^4。对于n个开关，可能的配置总数就是2^n。这和我们数有n位小数位的小数的数量时遇到的情况是一样的。只不过，现在每一层只有两个选项（0和1，或者说“开”和“关”）而不是10个选项。这和刚刚的菜单问题也是一样的。

二进制系统的工作方式就像是一系列开关，类似于十进制系统，只是这里的单位不再是个、十、百、千……，而是个、二、四、八……。当我们考虑所有整数时，二进制比十进制更加常用。比如，我们可以像下面这样将十进制四位数和二进制四位数做个对比：

	第一位	第二位	第三位	第四位
十进制	$\times10^3$	$\times10^2$	$\times10^1$	$\times1$
二进制	$\times2^3$	$\times2^2$	$\times2^1$	$\times1$

所以十进制数1101在十进制中的展开式是下面这样的：

$$(1\times1000)+(1\times100)+(1\times10)+1$$

而二进制数1101转换成十进制的展开式是下面这样的：

$$(1\times8)+(1\times4)+(1\times2)+1$$

相当于十进制数13。

在十进制中，一个四位数可以表示小于等于9999的任何数字。9999就是10^4-1。然而在二进制中，四位数可以表达的最大的数字仅仅是$2^4-1=15$。

这种说法让二进制显得没什么用。可是，正如我在第5章里讲的那样，使用二进制，我们可以用自己的手指数到1023。这是一种取舍。在二进制里，我们可以使用简单的开关编码任何事物。而在十进制里，每一个“位置”可以表达10种不同的信息，0、1……一直到9。在一些情况（比如电脑）中，你有足够的容量承载数字的位数，而每一位表达的状态非常有限。但是在另一些情况（比如书籍的ISBN编码）中，你能利用的空间位置非常有限，但是每个位置能承载的信息则非常多。HTML（超文本标记语言）颜色编码更加节省空间，它是用十六进制表示的。十六进制的基数是16，也就是说，每一个位置有16种可能的数字符号，这些数字符号分别是0、1、2、3、4、5、6、7、8、9、A、B、C、D、E、F。

一个我很喜欢的二进制应用方式是点生日蜡烛。假设我们有7根蜡烛，看上去这些蜡烛好像只能装饰不大于7岁的儿童的生日蛋糕。但是，如果使用二进制，用点燃的蜡烛表示“开”或者1，用熄灭的蜡烛表示“关”或者0，这个时候，这些蜡烛就足以装饰128岁以下的任何人的生日蛋糕了。换句话说，我们差不多可以给地球上的任何人过生日了。

通过二进制，我们可以用自己的手指一直数到1023，也就是$2^{10}-1$。这个说法是对的。下面，我来讲一讲我们是怎么做到的。每根手指都有两个可能的位置：伸开或者蜷起来。伸开表示1，蜷起来表示0。现在，我们就有了十位数的空间，可以用十位的二进制数表示从0到1023之间的每个数。图7–5显示了我们怎样用一只手表示从0到31之间的每一个数。



图7–5

每张小图所对应的五位二进制数如下所示：

00000	00001	00010	00011	00100	00101	00110	00111
01000	01001	01010	01011	01100	00101	01110	01111
10000	10001	10010	10011	10100	10101	10110	10111
11000	11001	11010	11011	11100	11101	11110	11111

二进制数很令人满意，但使用起来需要集中注意力，它不像直接用手指从1数到10那么直观。所以不幸的是，这个办法不能帮助我们节省脑力。我可以肯定，当我和别人交谈时，如果我用这个办法数数，那么我能数到的数字不会很大。我尝试过，一旦数字超过10，我就会开始犯错。

事实上，如果你足够专心而且能够很好地控制自己的手指的话，你甚至可以用手指表达三进制数字。这意味着，每个手指需要三个可能的位置。这需要你能将自己的每个手指蜷缩一半，而且不引起其他手指的联动。这样的话，每个手指就有三个可能的位置，分别表示0、1和2。尝试一下举起无名指，然后让中指自己蜷缩一半和完全蜷缩。（这通常有点儿困难。）

以上讲的都是关于整数的，我们同样可以使用二进制来表示分数。我们只

需要知道十分位、百分位的含义，并且把它们对应到二进制中的“二分位”和“四分位”上。另外，我们必须记住我们说的是二进制小数而不是普通小数。

当我们在处理普通的十进制小数时，小数点后面的位置分别代表$\frac{1}{10}$、$\frac{1}{100}$、$\frac{1}{1000}$……以此类推。比如下面这个数字：

$$0.3526$$

其真正的含义是：

$$\left(3\times\frac{1}{10}\right)+\left(5\times\frac{1}{100}\right)+\left(2\times\frac{1}{1000}\right)+\left(6\times\frac{1}{10000}\right)$$

在二进制当中，小数的第一位表示$\frac{1}{2}$，第二位表示$\frac{1}{4}$，第三位表示$\frac{1}{16}$。所以二进制数字：

$$0.1101$$

转换成十进制是：

$$\left(1\times\frac{1}{2}\right)+\left(1\times\frac{1}{4}\right)+\left(1\times\frac{1}{8}\right)+\left(1\times\frac{1}{16}\right)$$

也就是0.8125。我们可以用树状图来表示可能的二进制小数（见图7–6），类似于之前做的开关配置图。

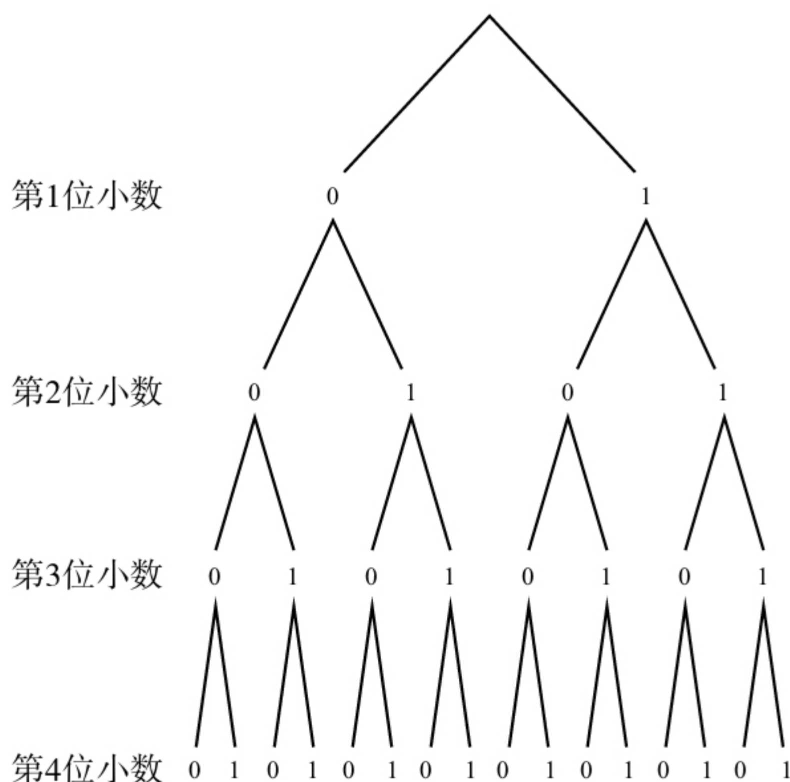


图7–6

在图7–6中，我们用0和1代替了之前的“关”和“开”。这有点儿像之前在做二进制生日蛋糕蜡烛时做的：用点燃的生日蜡烛表示1，用熄灭的生日蜡烛表示0。在这里我们处理的是有4位有效数字的小数，所以我们画的树状图有4个层级。我们也可以使用树状图来描述十进制小数，只是需要在每一个节点的下一层画上10个枝杈，但这样一来，我们的树状图很快就会没地方画了。

现在，每片叶子代表一个二进制小数。我们可以选定一片叶子，然后沿着从树顶到那片叶子的路径进行观察，记录下每一层的0和1，从而确定这个小数是什么。通过这种办法，我们可以确定第一片叶子是0.0000、第二片叶子是0.0001……以此类推。图7–7中标注了所有这些小数是什么。

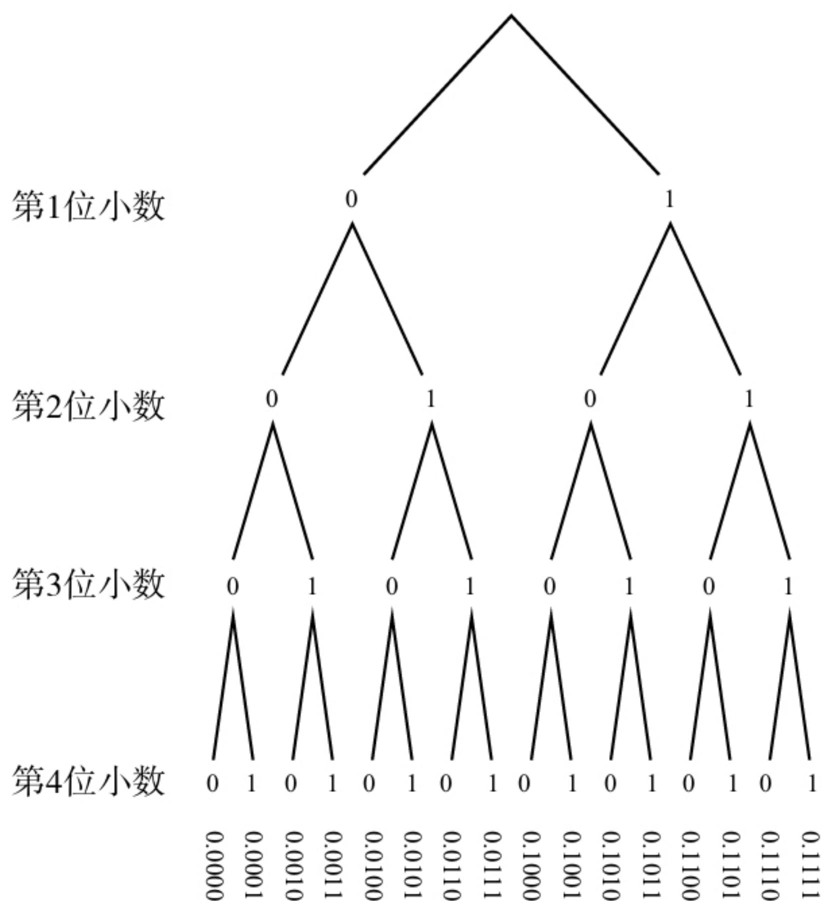


图7–7

要处理有5位有效数字的二进制小数，我们需要画一个有5个层级的树状图。要处理有n位的二进制小数，我们需要画一个有n个层级的树状图。要处理有无穷位的二进制小数，我们需要一个“有无穷个层级的树状图”。这个概念听起来有一点儿奇怪，因为最终的数字要通过这个树状图的最后一层，也就是树状图的叶子来表示。但是，如果这个树状图有无穷个层级的话，那么就不会有最后一个层级。我们在前面看到，我们可以选定一片叶子，然后从树顶出发，找到通往这片叶子的路径，进而确定这个数字是什么。但现在我们没有确切的终点了，因为树状图有无穷层。但我们还是可以通过树状图找到路径，只是这些路径是无穷长而没有终点的。

二进制系统可以表示十进制系统能够表示的所有数，只不过，二进制系统

通常会使用更多的位数。同样的，二进制小数可以表示十进制小数能够表示的所有小数，当然也会使用更多的小数位数。所以，无穷的树状图的路径能够对应所有的“无穷小数”，也就是所有的实数——或者至少能够对应从0到1这个区间内的所有实数。接下来，我们可以试着进行一下思维上的跳跃：

* n层的树状图有2^n条路径。

* 无穷层的树状图有2的无穷次方条路径。

我们还没有开始思考小数的整数部分，但是在下一章里，我们将会认识到为什么那部分并不重要。我们也会看到为什么使用二进制比使用十进制更加便利。

你可能认为到这里为止，我们依旧没有取得什么实质性进展。因为我们还是不能回答下面这个问题：2的无穷次方到底有多大？我们将在下一章里搭建通往这个问题的答案的桥梁。但是，首先让我们为自己到达现在所在的位置鼓鼓掌。我们已经学会了怎样使用树状图在自然数的基础上构建实数。我们也已经看到，这会帮助我们找到比无穷更大的无穷。我们需要做的就是重复这个过程，构建出一层又一层的无穷。这就是下一章的主题。

8　无穷vs无穷

当孩子们第一次学会爬一小步的时候，他们完全着了迷，会爬得越来越高、越来越高。他们惊叹于通过重复刚刚学会的一个小动作竟然能够爬那么高。如果他们能够爬上一个非常高的地方，他们就会特别兴奋。

我们刚刚就学会了怎样爬上一个非常高的地方——怎样从一个无穷往上爬到另一个更高（大）的无穷。就像是小孩子惊叹于他们刚刚学会的攀爬技能一样，我们也要一个台阶接着一个台阶地爬上更高的地方。这里说的并不是物理上的台阶，甚至不是数字上的台阶。我们将要爬的是无穷这个台阶。

我们将会利用到下面这些事实：

* 实数比自然数更无穷。

* 如果把自然数集的无穷写作ω的话，那么实数集的无穷应该写作2^{ω}。

我们将重复这个步骤，证明我们可以进行无穷的迭代，获得更加无穷的无穷。当然，我们首先要搞清楚这句话到底是什么意思。

通常的问题是：我们怎样才能比较无穷的大小？到目前为止，我们已经认定了实数是不可数的。换言之，我们没有办法精确地把它们和自然数对应起来，因为一些实数必将因此被落下。从直觉上看，这意味着实数的数量要比自然数“多”。但是如果两种类型的数字的数量都是无穷的话，一个比另外一个“多”到底是什么意思呢？一些无穷比另外一些无穷大——这怎么可能？无穷不已经是无穷大的了吗？无穷不是存在的最大的事物了吗？怎么可能还有事物比它还要大呢？

就像是灵魂、永生或者我是不是胖这些问题一样，答案的关键在于定义。比如，“胖”的定义到底是什么？在无穷这个问题上，症结在于“大”的定义是什么。我们可以直接放弃，然后说，这种讨论没有意义，因为无穷就是无穷大。但是数学从来不会放弃，除非得到一个逻辑上的悖论。如果一个现象在直觉上是合理的，那么数学家一定会去找一个框架证明它在逻辑上是合理的。我们数学家就是这么固执。如果我在徒步旅行的时候遇到一条路而这条路通往悬崖的话，我一定会回头。但是，如果我遇见一个数学上的悬崖，或者一条横亘在直觉和逻辑之间的鸿沟，我一定会急着走过去观察一下这个悬崖，看看我能做点儿什么。

这通常会牵扯一些细致的标准——是不是听起来有一点儿循环论证的意味？如果我想要证明我不胖，我可以选择体重指数作为标准。这样，根据最“权威”的定义，我并不胖。但是如果我想要证明我胖的话，我也可以选择使用腰围臀围比作为标准。这样，根据这一最常用的标准以及糖尿病的风险评估标准，我是胖的。

和人们对数学的传统理解比较起来，这种通过操纵标准来满足自己目的的做法听起来就像是一种倒退。根据人们对于数学的理解，数学家们总是先确定研究对象——比如数字，然后再研究关于数字，什么说法是正确的。比如，当你把3和4加在一起的时候会发生什么？当你把4和3加在一起的时候又会发生什么？啊哈，原来两次计算的结果是一样的！这样，我们就发现了一个关于数字相加的规律。

虽然你可能没有注意到，但数学并不总是这样的。比如，解方程就是一个例子。如果给你一个等式，比如：

$$3x+4=10$$

这相当于是在说：我想要这个等式成立，那么什么样的x能让这个等式成立呢？很多高等数学都始于数学家梦想让一些事物成立，然后开始构建一个能够让梦想成为现实的世界。我觉得在现实生活中，我们也是这样实现自己的梦想的。

现在，关于无穷我们希望实现的是：我想要找到比无穷更大的无穷。那么我要怎么定义“更大”才能让这个论断成为现实呢？我们的模板就是实数比自然数“多”这个事实，虽然实数和自然数的数量都是无穷的。我们可以用这个事实作为“更大的无穷”的定义，看看会发生什么。接下来发生的事情将相当有趣：我们要来看看如何一层一层地构建越来越大的无穷。

比较不同的集合

在上一章里，我们证明了实数是不可数的，证明方法是展示任何将实数和自然数一一对应起来的尝试，其都会导致至少落下一个实数的结果。我们可以用这种办法来比较任何两个事物的集合。对于非常小的集合，这个办法是有效的。

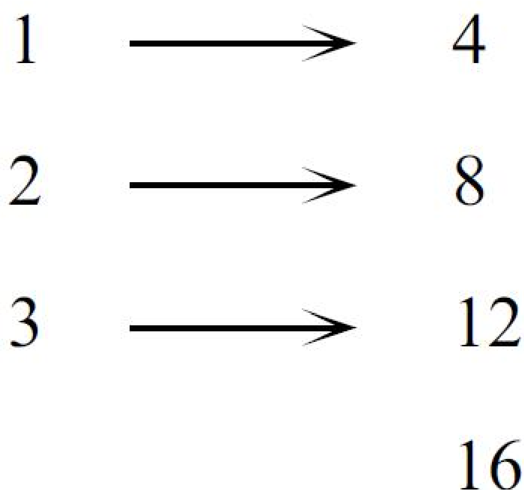


图8–1

我们在前面见过类似例子。在图8–1这个例子里，16被落下了。无论我们怎么努力配对，我们最终都会在右边剩下点儿什么，即便不是16，也会是其他数字。

如果将集合**A**和集合**B**一一对应的每一种尝试都必将失败，而右边总会剩下什么的话，那么我们就说，右边的集合“更大”。

技术层面的情况比这更加复杂一点儿。上面的不可能（每一种尝试都必将失败）说的是对于从左到右的满射函数来说，这是不可能的，但严格意义上讲，我们还要说明从左到右的内射函数是可能的。这个层面的问题非常微妙，涉及选择公理。在所有我们能想到的情况里，内射函数都是非常明显的一种情况。比如，从自然数到实数有一个非常显而易见的内射函数存在，因为我们可以把自然数直接和实数集合里的自然数部分对应起来。

从某种程度上讲，这有点儿像是小孩子在还不会数数的时候使用的最基本的比较他们有多少东西的办法。比如，如果要看看自己是不是有足够的饼干能分给朋友，他们可能要把饼干都分出去才能知道。但是会数数的人可能数一下就知道了。要检查我们在《舞动奇迹》里面是不是有同样多的明星和专业选手搭配，我们只要观察是不是每个人都搭配上了就可以了。如果是的话，那么两者的数量一定是相等的。我们并不需要数，因为我们并

不需要知道一共有多少个组合。但是，即便我们真的去数的话，我们在真正意义上做的也不过是把每一个组里的人和数字的官方袋子做对应而已。假设一共有10个挑战者。当你数他们的时候，你真正做的就是把他们和下面的数集做对应：

{1、2、3、4、5、6、7、8、9、10}

如果你把他们和这个数集里面的数字做了对应，然后发现你没有用到9和10，那么你就会知道，挑战者的人数不到10。如果你想要对应的时候发现你用了9和10两次，那么你就会知道，挑战者的人数超过了10。

对于这种小的数字，使用这种分析方法确实有点儿小题大做了。但这是理解数学的常见方法。你需要先在简单的例子中理解这个方法，如果你一上来就在非常复杂的情形下学习这个方法的话，那么学起来就太难了。

无论如何，这种对应方法是小孩子在学会数数之前就可能会用到的方法。他们也许有一列乐高火车，而且他们想要在每一节车厢里放一个乐高小人儿。如果乐高小人儿没有了但是还有空的车厢，那么很明显，他们没有足够多的乐高小人儿，虽然他们并不知道他们有多少乐高小人儿，也不知道他们还需要多少乐高小人儿。这就是我们在处理无穷数集的时候，用来处理“更多”和“更少”的问题的办法。就像是不会数数的小孩子一样，当我们拥有无穷个事物的时候，我们并不确切知道我们到底有“多少”。

让我们假设我们有无穷多个乐高小人儿，要放到一列有无穷节车厢的乐高火车里。我们想要知道，乐高小人儿的数量是不是比火车车厢的数量“更加无穷”。现在，你可能会想，如果我们最后还有空的车厢的话，那么我们肯定没有足够多的乐高小人儿。我们在这里面临的第一个问题就是，我们不知道“最后”是什么意思。如果我们要把无穷个乐高小人儿放到无穷节的车厢里，我们可能要花费“无穷”的时间——无休无止。但我们可以忽略这个事实。因为我们只需要想到一个做这件事的办法，而并不需要真的做这件事。

这是一个我们能够把所有的小人儿放进车厢并且剩下一个空车厢的做法：我们可以空下第一节车厢。也许乐高小人儿像我一样不愿意坐在第一节车厢里，以免在发生撞车事故时受伤。所以我们把第一个小人儿放在第二节车厢，下一个小人儿放在第三节车厢，以此类推。最后，在我们填充了所有的车厢之后，我们还是有一节空的车厢。但是，这并不意味我们没有足够多的乐高小人儿。我们可以把每一个乐高小人儿都往前挪一个车厢，这样所有的车厢就都满了。

我们甚至可在在第一次放小人儿的时候空出无穷多节车厢。我们可以只把

乐高小人儿放到偶数号的车厢里，即把它们放到2号车厢、4号车厢、6号车厢……就像是转移希尔伯特旅馆里的客人时做的一样。这样，所有奇数号的车厢就都是空的了。看起来好像我们没有足够多的小人儿，但事实上，我们的小人儿数量是足够多的。

这是探索无穷的过程中发生的众多奇怪而美妙的事情之一。这意味着，我们要么放弃，要么就要对我们表达的意思格外小心。我们的无穷数集可以被安排得非常奇怪，一些安排能够让乐高小人儿占满所有的车厢，而另一些安排则不能。这就和把更多的人安排进希尔伯特旅馆时发生的事情那样。只有10个小人儿和10节火车车厢的时候就不会发生这样的事情。如果你想要在每一节车厢里都放上一个小人儿，你可以使用很多不同的办法。但是只要你用尽了你的小人儿，你就一定填满了所有的车厢。如果你的小人儿和车厢的数目是其他的有限的数字，并且你用尽了小人儿但是仍然有空的车厢，那么你就能确定地说，车厢的数量比小人儿的数量多。

在拥有无穷多的小人儿和无穷多的车厢的情况下，你有可能会用尽你的小人儿而还有剩下的车厢，但你不能确定你是不是能够通过重新安排来填满所有的车厢。如果你真的想要证明车厢的数量比小人儿的数量多，你就要保证无论怎样安排，你都没办法让你的小人儿填满你所有的车厢。这看起来很困难，但是事实上，我们曾经用对角论证法做过一次了。那个时候，我们证明了实数和自然数之间并不存在完美映射函数。如果存在这样的函数就会出现矛盾，因为至少会有一个实数被落在外面。

在数学领域里，我们不会在这种情况下使用“更大”这个词。因为这个词太模棱两可了。如果我们要用“更大”这个词的话，我们就必须非常小心地定义“大的程度”这个概念。在数学里，这个概念叫作“基数”。一个集合的基数就是这个集合中的物体的数量。如果一个集合所包含的物体的数量是有限的，那么这个集合的基数就是其中物体的数量。如果一个集合所包含的物体的数量是无穷的，问题就会变得复杂一些。现在，我们的状态很像一个还没有学会数数的小孩子。虽然我们并不知道集合的基数到底是什么，但是我们已经知道怎么描述比较的结果：一个集合的基数和另外一个集合的基数一样大，或者一个集合的基数比另外一个集合的基数大。据此，我们可以开始搭建基数越来越大的无穷集合。

最小的无穷

最小的数集是空集。所以最小的基数就是0。除此之外，基数还有可能是所有的有限的数字。比如，只包含1个物体的集合、包含2个物体的集合、包含n个物体的集合。

事实上，自然数是最小的无穷数集。记住，我们这里说的“最小”这个概念

来源于我们对于大小的衡量，而大小的定义来自我们所使用的将事物做对应的方法。如果存在一个比自然数“更小”的无穷集合，那么它的含义是什么呢？其含义就是存在满足下列条件的无穷集合：如果我们想要把这个集合和自然数对应，我们就不可避免地会剩下一些自然数。

让我们再一次回到转移希尔伯特旅馆客人的例子来思考一下这个问题。一个比自然数“更小”的无穷集合意味着你有无穷多的客人，但是当你将他们转移到一个仅有一层的希尔伯特旅馆的时候，你总是会浪费一些房间。但这是不可能的，我们总是能够重新安排这些客人，让他们按照顺序搬进新的房间而不空下任何一个房间。这样说起来有一点儿含糊，但是我们可以给出非常精确的转移说明。比如，“数一数比你的房间号小的客人有多少，用这个数字加上1就是你的新房间的号码”。

✱原本房间号最小的那个客人前面有0个房间号比他更小的客人，所以这个客人数出来的数字是0。因此，这个客人将会搬到0＋1＝1号房间。

✱原本房间号第二小的客人前面有一个房间号比他更小的客人，所以这个客人数出来的数字是1。因此，这个客人将会搬到1＋1＝2号房间。

这不是一个完整的证明过程，但是这给了我们一个能够证明的观点。那就是每一个自然数的无穷子集都能够通过某种方式和自然数集对应起来。这就是说，自然数的子集要么是有限的，要么和自然数具有同样大小的基数。不存在基数介于两者之间的无穷数集。所以我们知道了，最小的无穷数集的大小，就是自然数集的大小。

还记得吗，在我们试图检验无穷是不是一个数字的时候，我们不停地遇到一个问题：从等式两边同时减去无穷会导致等式不成立。我们现在可以试着弄懂其中的原因了。从一个自然数的无穷数集中移除一个无穷子集的办法太多了。你可以把所有的数字都移除，什么都不剩下。你可以移除所有的偶数，留下无穷多个奇数。或者你可以移除所有大于10的数，剩下10个数。或者移除所有大于n的数，剩下n个数。也就是说，减去无穷可能得到各种各样的答案。我们在下一章会继续讨论这个问题。

鉴于现在我们正在构建各种不同大小的无穷，我们需要思考一下怎么为无穷找到更好的标记。当我们认真考虑不同大小的无穷的时候，我们会使用$\aleph_0$这个符号，它读作“aleph”，是希伯来语字母表中的第一个字母；下标0表示这个数集是存在的最小的无穷数集，也是无穷的等级系统的最初一级。

下一个无穷

在论证中，我们发现了很多无穷其大小都和自然数一样。偶数或者奇数这两个无穷看起来好像小一点儿，其实也并不是。我们甚至可以更加进一步，只考虑100的倍数，或者100万的倍数。后面这个无穷集合只包含自然数数量的1/1000000的数，但是它的基数还是和自然数集一样大。后面这个论证为下面这句话提供了事实依据：“无穷除以100万还是无穷”。我们可以只考虑大于100的数。但是这个数集的大小还是一样的。这个论述为下面这句话提供了事实依据：“无穷减去100还是无穷”。

我们也遇到过一些看起来更大一点儿的无穷，但事实上它还是和自然数一样大。比如我们把两个自然数集加起来——红色的自然数集加上蓝色的自然数集。我们也可以复制三个自然数集，或者复制可数的无穷个自然数集。或者我们可以使用整数集或者有理数集来复制。但是我们会发现这些无穷数集还是能和自然数集对应起来。这说明这些无穷数集仍旧是$\aleph_0$。

到目前为止，我们只见过一个事实上更大一点儿的无穷数集，就是实数集。问题在于，它是比自然数集稍大一点儿的下一个数集吗？还有没有大小介于二者之间的无穷数集？这是一个非常难的问题，而连续统假设（Continuum Hypothesis）解决了这个问题。“连续统”在这个问题里指的是实数，因为人们认为实数“连续”覆盖了整个数轴。相对的，整数和自然数都在数轴上留下了空隙。这个理论被称为假设是因为至今为止这个假设还没有得到证明。这个假设是康托尔在1878年提出的。他提出实数的基数和自然数的基数之间没有其他的基数。换言之，没有哪个数集其基数介于自然数的基数和实数的基数之间。这意味着一旦你摆脱了自然数集，获得了比自然数集更大的无穷数集，那么这个数集就会立即变得像实数数集一样大。这个论述是否正确还没有得到证明。这句话在一些世界中是正确的，但是在另外一些世界中是错误的，它是否正确完全取决于你身处哪一个世界。

已经得到证明的事实是，在标准的策梅洛–弗兰克尔逻辑系统里，连续统假设是无法得到证明的。但同样得到证明的事实是，这个假设也是不能被推翻的。这意味着，我们可以找到另一个世界，在其中连续统假设是正确的。但是，我们也可以再找到一个世界，在其中连续统假设是错误的。这个结果并不取决于逻辑规则。连续统假设不能被推翻的证明是库尔特·哥德尔在1940年做出的，而连续统假设不能被证明的证明是保罗·寇恩在1963年做出的。这个发现极其重要，寇恩也因此获得了菲尔兹奖。

紧邻$\aleph_0$且比它大的那个无穷应该叫作$\aleph_1$，所以表达连续统假设的另一个方法是说实数的基数是$\aleph_1$。我们可以做一些进一步的推导。在上一章，我们根据自然数的基数推导出了实数的基数是什么：如果自然数的基数是$\aleph_0$的话，那么在二进制的基础上实数的基数就是$2\aleph_0$。

你可能注意到了，如果我们使用十进制的话，实数的基数就变成了$10^{\aleph_0}$。在这里，使用二进制更好的理由有以下几个。首先，10是一个非常人为的选择。它之所以能够在实数计算中被大众使用这么久，仅仅是因为我们有10个手指，因而更偏爱十进制的数字。然而，2并不是人为选择的数字。它是最小的进制意义上的基数。

使用二进制更合适的另外一个原因是$2^{\aleph_0}$看上去比$10^{\aleph_0}$小一点儿。我们现在想要做的事情之一就是看看这个数字是不是那个大于$\aleph_0$且和$\aleph_0$最接近的无穷数集。所以，让这个数字的表达式看起来小一点儿似乎更符合要求。事实上，两种表达方式并没有差别，因为$2^{\aleph_0}$和$10^{\aleph_0}$其实一样大。实际上，这两个数字都和$1000000000000^{\aleph_0}$一样大。但是如果我告诉你$1000000000000^{\aleph_0}$是大于$\aleph_0$且和$\aleph_0$最接近的无穷数集的话，你很可能会觉得1000000000000这部分错了。另外，把这个数表达成$2^{\aleph_0}$确实更加具有启发性。这告诉我们，为了构建一个更大的无穷，我们需要把前一个无穷放到指数的部分，即便它的底数像2这么小也没问题。

你现在可能想起来我们一直忽略了小数点前面的部分。这是我比较不在乎的地方，因为这件事并不会造成什么不同。我们可以试着修复一下。小数点前面的部分是整数部分。而我们知道一共有多少整数——有$\aleph_0$那么多。这意味着我们需要将我们计算出来的实数数量乘以$\aleph_0$来矫正我们对于整数数量的忽略。然而，用这个可数的无穷来乘巨大的$2^{\aleph_0}$并不会让后者变得更大。实际上，一直以来，我都暗暗知道为什么我不担心小数点前面的部分。

从几何学的角度讲，如果我们忽略整数部分，那么我们实际上只是在看0到1这个区间内的实数。如果我们在实数轴上看这一小段线段，我们需要复制多少次才能获得整个实数轴呢？每有一个自然数，我们就需要复制一次，所以这段线段需要被复制$\aleph_0$次。另外，我们可以把注意力集中在实数轴上的任意一段线段上。无论这段线段有多短，我们从上面获得的实数数量都是一样多的。理解这个事实的一个方法就是构建一个双射函数，也就是构建从这段线段上的数字到0和1之间的数字的完美映射。比如，我们想要看看0和$\frac{1}{100}$之间的实数。我们想要把这些实数对应到0和1之间的实数上。我们要做的事情很简单，就是把0和$\frac{1}{100}$之间的实数都乘以100，然后把这些实数和计算出来的数字对应起来。这有点儿像把小线段按比例放大成大线段（见图8–2）。

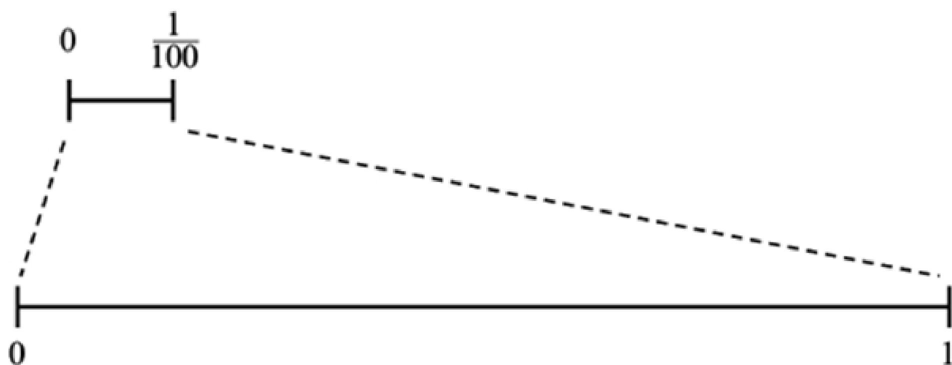


图8–2

这个论述对于任何大小的区间都是成立的。换言之，从任何实数开始到任何其他实数之间的实数数集都满足这个论断。不过，要把0到1这个区间内的实数放大成整个实数轴，我们需要更狡猾一点儿。因为我们不能只是“把一切数字和无穷相乘”。我们可以用下面的方法获得所有的实数：假

设x是你从0到1这个区间内找到的任意一个实数。你可以计算$\frac{1}{x}$，这样我们

就获得了一个位于1到无穷大这个区间内的实数。然后减去1。这样我们就把0到1这个区间内的实数和0到无穷大这个区间内的实数对应起来了。

现在，我们知道了实数集的基数是什么，我们可以用下面的公式重新表述连续统假设：

$$\aleph_1 = 2^{\aleph_0}$$

我们已经学会了怎么在无穷的阶梯上往上爬一级。如果你像小孩子一样对数学充满好奇的话，那么你就会想继续向上爬。比如，把这个新的无穷作为指数，把2作为底数得到一个新的幂。连续统假设的广义版本说的是，在任何水平上，用这种方式找到的无穷都是比原来的无穷更大且最接近原来的无穷的新的无穷。通过这种方式，我们能够搭建出无穷的层级。

$$\aleph_0、\aleph_1、\aleph_2、\aleph_3……$$

在每一个水平上，都有$\aleph_{n+1} = 2^{\aleph_n}$。因为还存在更加晦涩版本的连续统假设，所以我们无法证明这就是获得无穷不同层级之间距离最接近的无穷队列的方式。但是至少我们确实看见了无穷正在变得越来越大。

这些无穷真的变大了吗？

如果$\aleph_0$是“自然数集的大小”，那么我们为什么能把它作为2的指数呢？我们通常定义2^n的方式是“2和自身相乘n次”，但是在无穷的情况下，这个说法是不成立的，因为你不可能让“2和自身相乘无穷次”。

这里的关键是我们不能继续用对待数字的方式对待2^n，并牢记应当把基数看作一个数集的大小。如果你愿意的话，你还可以把这些数集想象成数字袋子。所以n实际上是一个包含n个元素的集合的大小。我们把2^n想象成另外一个集合的大小，而这个集合在某种程度上和刚刚那个有n个元素的集合相关。提出这个观点的目的是把一些对于有限数字来说有道理的方法应用到无穷集合上。因为我们知道，$\aleph_0$的本质不是其他事物，而是自然数集的大小。

我们接下来要讲的事情就和你选择带什么鞋子去度假差不多。当你决定这是一件重要的事情的时候，你可以盯着你的鞋子，然后挑出你愿意带上的所有鞋子，或者你可以一双一双地检查，再依次决定带不带它。所以，你可以选择带上的鞋子一共有多少双呢？如果你一共只有两双鞋，一双是运动鞋，另外一双是凉鞋，那么你就有以下几种选择：

* 什么鞋都不带，光脚去。

* 只带凉鞋。

* 只带运动鞋。

* 既带凉鞋，也带运动鞋。

一共有4种可能性。另一种方法是针对每双鞋子决定是否带它。我们画了一个决策树来表示这种方法（见图8–3）。

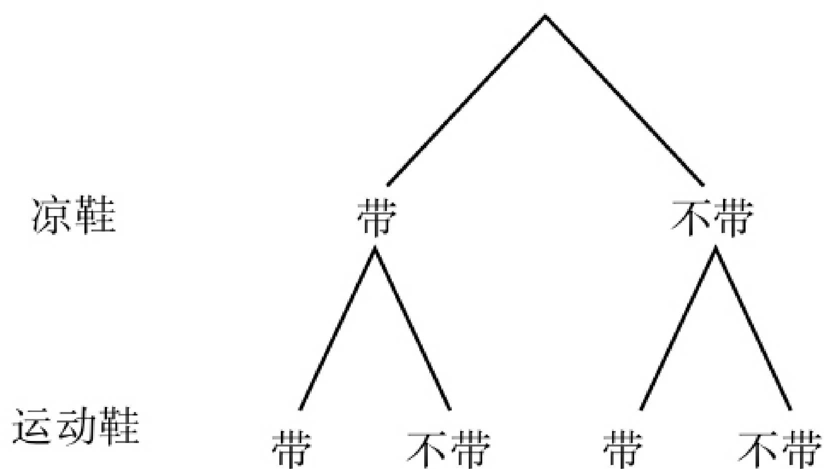


图8–3

是的，如果你真的画出这样的决策树来帮助自己决定是不是带你的两双鞋子的话，别人大概会觉得你非常荒谬并嘲笑你吧。许多数学家就会时不时做出这种举动。有一次，在为一个大型的聚会准备食物的时候，我就曾经画了一个流程图来缩减花在厨房的时间。当然，我在别人看见之前就尽量把这些图纸扔掉了。

你可能想起来了，我们曾经见过这个树状图。我们展示菜单选项和二进制分数的时候都用过这个树状图。如果我们有三双可供选择的鞋子的话，这个树状图就会变成图8–4的样子：

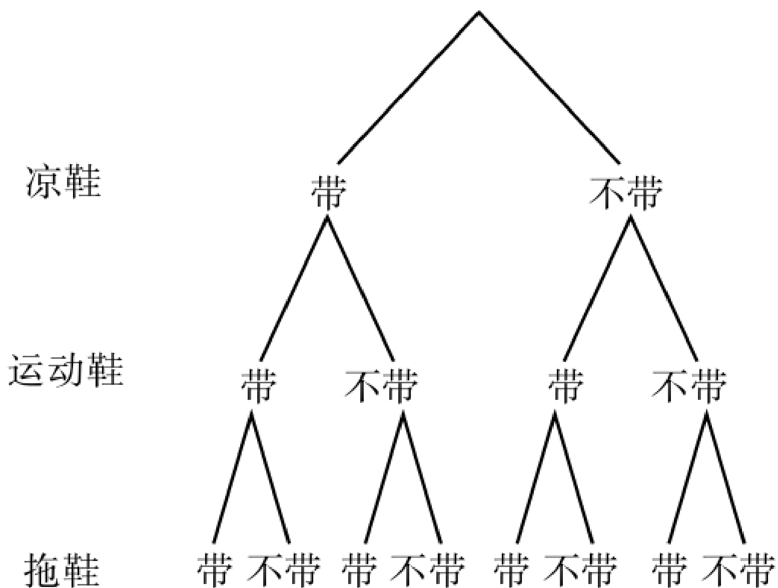


图8–4

所以，现在可能性的总数是$2\times2\times2=2^3=8$。

如果我们有n双鞋子可供选择，那么可能性的总数就是2^n，就像有n位的二进制数字的总数一样。

现在，我们需要做的是对于n为无穷的情况给出一个合理的解释。这实际上叫作归纳法，它是数学思考中的一个重要手段，指的是你从一些相当熟悉的事物开始，试着把一个对于熟悉事物来说合理的概念运用到一个不那么熟悉的事物上去而又尽量不毁掉原来的概念。之前，我们含糊地讨论了“无穷树”并且凭着信念向前探索了一步。而现在，我们将要做一些更加精确的事情。

从挑选鞋子带去旅行那里，我们弄明白了2^n这个概念。这意味着我们从我们的鞋子这个集合中挑选了一个子集。我们的结论是，如果我们一共有n双鞋子可供选择，那么我们挑选出来带去旅行的那个鞋子的子集的可能性就有2^n种。

现在，即便n等于无穷，这个概念也是成立的。换言之，如果n是我们构建出来的无穷的话，那么这个概念也还是成立的。我们从一个无穷的数集开

始，比如自然数集。根据我们的定义，自然数集的基数是$\aleph_0$。然后，我们可以开始考虑“包含自然数集的所有可能的子集”这个数集。接着，我们把$2^{\aleph_0}$“定义”为这个数集的基数。我们还不知道这个数到底是多少，但是我们已经为这个数赋予了意义。我们为这个数下定义的方式和给2的有限次幂下定义的方式是一致的。

我们不可能把所有这些自然数集的子集写下来。因为首先我们就不可能把所有的自然数写下来。事实上，现在的情形更加糟糕了。因为自然数集的子集是不可数的。我们没有办法把这些子集排成一列而不把某一个漏在外面，即便从理论的角度上讲也是不可能的。

如果你还记得$2^{\aleph_0}$是实数集的大小，你可能会觉得通过“自然数集的所有可能的子集的数量”来研究“实数的数量”这个做法非常随便。以免你为此感到烦恼，我们来看一看这些概念是怎样相互联系起来的。记住，为了挑选自然数集的子集，你有下面几种办法可以用：

- 列一张清单，写出包含在你的子集里面的所有的自然数。

- 或者，每次检查一个自然数，然后在它的旁边写上“包含”或者“不包含”来表示你的子集中是否包含这个自然数。

如果你使用的是第二种方法，你实际上做的事情就是写出一长串的“是”和“否”。而你可以进一步把这个文字串转换成由1和0构成的无穷长的数字串，而这正是二进制分数。

事实证明，我们可以使用对角论证法来证明一个集合的所有子集所构成的集合总是比原来的集合大。对于有限的数字来说，这个论证是显而易见的。无论你有多少双鞋子，你带去度假的鞋子组合的可能性的数量永远比你鞋子本身的数量多。对于无穷的集合来说，这个结论有点儿晦涩，但还是成立的。因为我们就是这样定义2^n的，而我们知道，2^n总是大于n。这意味着，我们提出的无穷的层级具有这样一个特点：

$$\aleph_0$$

$$\aleph_1 = 2^{\aleph_0}$$

$$\aleph_2 = 2^{\aleph_1} = 2^{2^{\aleph_0}}$$

$$\aleph_3 = 2^{\aleph_2} = 2^{2^{\aleph_1}} = 2^{2^{2^{\aleph_0}}}$$

那就是，它确实是在不断变大的。所以，无论我们能说出的无穷有多大，当我们想到它可能的子集的数量时，我们总是能获得一个更大的无穷。

这告诉我们，当孩子们说他们有无穷正确的时候，他们可以使用这个方法让他们的无穷比小伙伴的更大。他们可以说，“我的正确程度是2的无穷次幂！”然后就是，

“我的正确程度是2的2的无穷次幂次幂！”

“我的正确程度是2的2的2的无穷次幂次幂次幂！”

……

9 无穷是什么？

登山的路线有时是非常明确的。我不是一个职业登山家，所以我选择的徒步登山路线通常都是常有人走的路，路边甚至还会立着指示牌。然而，有一次我参加了一个法国爱丁堡公爵远征重游纪念活动，途中我经历了一段难忘的时光。当时，我们在离露营地不远的地方迷了路。实际的路程比地图上标注的要长，而且地图上的指南针方位和所有路径都不能完全对上。我们很累。我记得，我们那天一共走了42公里，背上还背着10多公斤重的双肩包（那个时候的帐篷非常重）。我们不想走错路再返回来，所以我们比较了不同的线索，看看哪一条路看起来最常有人走，哪一条路和地图对应得最好，哪一条路看起来能够通往一片空阔的空地。最终，我们在岔口处发现了一个用鹅卵石铺设的箭头。这一定是比我们先到这里的好心人留下的。

在本书的第一部分，我们走进了很多死胡同又转回身来，并在最终找到了一条有前景的通往无穷的道路。我们现在对于无穷到底是什么这个问题有了一些基本的线索。

* 数字可以用来测量官方的数字袋子（也就是数集）的大小。无穷也同样如此。

* 最小的无穷数集与自然数集的大小一样。

* 实数集这个无穷数集要比自然数集这个无穷数集大。

* 无论你现在想到的无穷是什么，你都可以用2做底数，用你想到的那个无穷做指数，从而构建出一个更大的无穷。

第一点好像是在说，无穷事实上确实是某种数字，虽然这跟我们在前面几章看到的不太一样，但只要我们用一种特定的方式看待数字，它就是成立的。整数、有理数和实数都是自然数的延伸，但是并没有在无穷方面为我们帮太大的忙。使用“数字的官方袋子”这种思考方式帮助我们拓展了对于自然数的理解，并且确实帮助我们初步理解了无穷。事实上，它为我们带来了两种不同的研究无穷的路径：一种是序数，另外一种是基数。我们将会研究一下这两种数字分别是什么，然后探讨这两种数字类型是怎样避开前面在我们试图将无穷定义为一个数字时遇到的问题。

我们前面讨论过基数就是测量一个数字“大小”的数。具体而言，基数就是我们用来测量不同数字的大小的官方袋子，特别是可以被用来测量越来越大的无穷。有限的基数和有限的自然数在实际应用中并没有什么差别。

但是，我们还需要关注另外一个研究无穷的角度。让我们回到我们第一次讨论的希尔伯特旅馆的问题：我们的旅馆已经住满了，但是现在又来了一位新的客人。安排这位新的客人入住看起来好像是不可能的，但是我们随后想到了这个办法，即把每一位已经入住的客人往后挪一个房间。这听起来好像没有什么问题，但实际上，我们完全忽略了我们给所有其他的客人造成的麻烦。如果这个单独的客人是先来的而无穷的客人是后来的，那么所有人就都不需要换房间了。换言之，这是一个不会带来麻烦的方案。如果无穷的客人先到而单独的客人后到，那么所有先到的客人就必须向后挪一个房间来为后到的客人腾出房间。换言之，这是一个必将牵涉一些“麻烦”的方案。

这就是客人到达顺序的问题。你可以想象，1号房间是最好的、2号房间是次好的、3号房间更差一点儿，以此类推。如果是这种情况的话，把后到的客人安排到最好的房间看上去有点儿不太公平。而你通常希望能够按照到达的顺序安排客人。

这让我想起有一次我在皇家阿尔伯特音乐厅排队参加舞会的事。每个人到达的时候都会拿到一张票，上面写着你在队伍里的位置。有一次我早上6点就到了，就是为了排队看由古斯塔夫·杜达梅尔指挥的西蒙·玻利瓦尔管弦乐队表演《马勒第二交响曲》。这是我最喜欢的曲目之一。我感到非常兴奋，因为我排在队伍的第6号。而在当了一整天的第6号以后，我和第5号成了相当好的朋友（第7号和第8号则是我的父母）。在这种情形中，人们到达的次序变得非常重要。我之所以那么早去排队完全是因为我真的想要排在前面。在排了13个小时的队之后，我非常不愿意任何人插队到我前面去。而排在第1、第2和第3号的观众都是通宵在馆外搭帐篷排队才拿到这些位置的。

这和你在酒吧里面举行一个聚会不同。如果房间最多只能容纳100个人，你可以在进门的时候给每个人一张票。这样，如果你给出了100张票，你就知道里面已经满了。你不在乎人们到达的顺序，你只需要知道你是不是放进去了100个人。

这就是基数和序数之间的差异。基数只测量事物的大小，而不关心这些事物彼此之间的相互关系。而序数会考虑事物的次序。这就像是希尔伯特旅馆的经理遇到的问题一样，为了让新到的客人入住而让其他的客人挪来挪去，这为大家带来了麻烦。

如果只考虑有限的数字，我们一般注意不到这两类数字之间有明显的差异。但是，如果我们思考的是无穷的数字的话，两者就会变得非常不同——序数比基数微妙得多。

在前面一章里，我们讨论了事物的绝对大小，并且把这个大小称之为“基数”。基数和序数之间的一个重要差别就是它们中包含多少种不同的无穷。我们看到，从大小的角度讲，最小的无穷就是自然数集，接下来是实数集。然而，如果我们把人们到达的次序和让希尔伯特旅馆里的客人们搬来搬去带来的麻烦纳入思考范围的话，那么在自然数集和实数集这两个无穷之间，一定还存在序数的无穷。如果你在乎客人到达的次序，情形就变得非常不同，也更加奇怪了。

希尔伯特旅馆安排新到达的客人的例子告诉我们：“1加上无穷还是无穷”不会造成任何麻烦或者不公平，但是“无穷加上1还是无穷”就会造成麻烦。我们把这个事实写成下面的形式：

$$\infty + 1 \neq 1 + \infty$$

现在，让我们再考虑一下两辆分别装着无穷多位客人的巴士。如果你让第一个巴士上的客人按照顺序入住旅馆的话，当第二辆巴士到达的时候，你就会遇到问题。你可以让所有先到的客人搬到他们的房间号乘以二的房间中，然后把所有的奇数号房间空出来。但这样是不公平的，因为后到的客人会在先到的客人还在找自己房间的时候先住到空出来的房间里去。我们在这里遇到的问题是：“无穷乘以2还是无穷”会带来麻烦。

现在换一种假设，这次不是有两辆分别装着无穷多位客人的巴士，而是有无穷的两人组。换言之，你看到无数辆双人自行车依次到达，每辆自行车上都有两位客人。这种情况可能比上一种情况现实得多。

第一辆双人自行车到达了，你让坐在前面的人去了1号房间，让坐在后面的人去了2号房间。

接着，第二辆双人自行车到达了，你让坐在前面的人去了3号房间，让坐在后面的人去了4号房间。以此类推（见图9–1）。

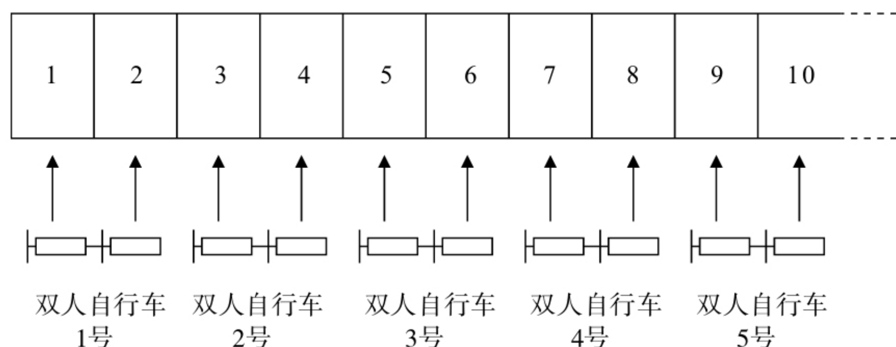


图9–1

每当有一辆新的双人自行车到达的时候，你都要把新到的这两位客人安排进后面的两个房间。这里的关键在于，没有人需要挪房间。在前面一个例子里，我们用“无穷乘以2”来表示一辆装着无穷位客人的车先到，另外一辆装着无穷位客人的车后到。这里，我们用“2乘以无穷”来表示有无穷组成对的客人依次到达。通常在数字都是有限的情况下，我们并不需要区分这两者。比如，我们并不需要区分“2×5”和“5×2”。因为我们知道这两种算法的结果是一样的。但是在这里，我们发现“无穷乘以2”会给旅馆带来麻烦，但是“2乘以无穷”可以在不带来麻烦的情况下让所有的客人入住。也就是说，如果我们把为客人带来的麻烦纳入考虑范围的话，我们面对的就是这样的情况：

$$\infty \times 2 \neq 2 \times \infty$$

这是另外一个线索，它告诉我们，如果我们考虑麻烦的话，无穷和常规的数字是不一样的。

你可能会觉得，“麻烦”这个概念不是很符合数学的理念。确实如此。这就是为什么我们需要从次序的角度考虑这个问题。你可以检查入住旅馆的客人的房间号是不是按照他们到达旅馆的次序安排的，这样，你就能检测到是不是发生了“麻烦”。如果你不在乎为客人们带来的麻烦的话，你就可以忽略他们到达的顺序。这种情况下，你只需要关注你是不是把每一位客人都安排入住了。这被称作“基数算术”，因为基数算术只在乎一共有多少事物。但如果你确实关心人们到达的次序的话，那么你用到的数学类型就是“序数算术”。序数算术不但关心一共有多少事物，还关心安排事物的次序。当我们考虑有限事物的时候，基数算术和序数算术是一样的。但是，在处理涉及无穷的问题时，两者之间就会出现有趣的差别，让我们认识到存在两种不同的但同样有效的无穷。

无穷的队列

由于现在我们开始考虑人们到达的次序了，因此有必要考虑一下队列的问题。想象一下你正在发放票据，以此标记人们在队列中的次序。这将会是一个非常非常长的队列，一种“希尔伯特队列”。因为队列里的人是无穷的。你手里有一整本票据，每张票据对应一个自然数，编号有1、2、3、4、5、6……，当人们到达的时候，你会按照他们到达的次序给他们发一个编号。

现在，假设你用光了所有的票据，因为已经有无穷多的客人到达了，就和希尔伯特旅馆客满的情形一样。而这时如果有一位新客人到了，你可以让每一位客人都往后挪一位，将1号票据拿回来，然后你就可以把这张票据给这位新到的客人了。但是这样做的话，这位新到的客人就插了所有人的队，这是非常不公平的。这时候，唯一公平的事情就是拿出一本新的其他颜色的票据来。假设第一本票据是红色的，第二本票据是蓝色的，现在你就需要记住所有拿着红色票据的客人都排在拿着蓝色票据的客人的前面。

你也许会想，有没有什么办法能够在不动用第二本票据的情况下解决这个问题？答案是，如果不允许插队的话，就没有任何其他办法了。这告诉我们，按照序数的理论，“无穷加上1”事实上比无穷要大。当我们用序数而非基数的方法思考自然数集这个无穷数集的时候，我们通常把自然数集写作ω。我们在这里总结出来的结果是$\omega+1>\omega$。相应的，$1+\omega$在这个例子里表示有一个人先到，然后无穷个人后到。在这种情况下，你可以轻松地只使用一本票据而用不着第二本票据，也就是说：$1+\omega=\omega$。

奇怪而有趣的序数算术才刚刚开始。在序数算术里，我们在常规数字中总结出来的原则都需要经过重新思考。发生改变的第一个原则就是数字相加的顺序。通常在有限的数字中，我们知道加法的顺序对加法的结果是没有影响的。所以才会有：

$$5+3=3+5$$

或者更加概括性地说：

$$a+b=b+a$$

这个等式对于所有的实数a和实数b都成立。需要记住的是，实数包括所有的小数，既包含有理数也包含无理数。我们已经看到了一个关于无穷序数的例子，在这个例子里上面的公式是不成立的。因为，虽然：

$$1+\omega=\omega$$

但是，

$$\omega+1\neq\omega$$

所以，

$$1+\omega\neq\omega+1$$

$a+b=b+a$这个特性（在它成立的情况下）被称作加法交换律，因为a和b是可以交换位置的。即便我们考虑序数的逻辑，这个规律在数字有限的情况下也是成立的。让我们来试一试5和3。

再一次假设你负责为队列发放编码票据。首先，5位客人出现了，你会给他们发放1号、2号、3号、4号和5号票据。这时，另外3位客人出现了，你会给他们发放6号、7号、8号票据。这就意味着，在考虑序数的情况下，$5+3=8$。

现在，如果那3位客人是先出现的，你会给他们发放1号、2号和3号票据。然后另外5个人出现了，你会给他们发放4号、5号、6号、7号和8号票据。结果仍旧是8。你并不一定需要按照这个次序给他们这些票据，但事实是，在这两种情况下你都给出了1~8号票据，而且没有发生插队的情况。所以，无论你用什么次序做这件事，结果都是一样的。

如果你觉得有一点儿奇怪，如果你发放的票据是2号、4号、6号、8号、10号、12号、14号和16号，或者是1号、5号、9号、14号、18号、100号、200号和378号，这也是一样的吗？一样。因为人们还是会按照他们到达的顺序排在队伍里。你只要记住，不要给出你略过的票据让其他人插队就可以了。

在1和ω的情况里，如果第一种情况是ω人先到，1个人后到，而第二种情况是1个人先到，ω人后到的话，我们是没有办法在两种情况下使用同样的票据而避免发生插队的情况的。你也许注意到，这并不是一个数学的论证方式，而仅仅是一个貌似合理的想法。要用真正数字的方法证明$1+\omega$和$\omega+1$是不同的，不妨想一想队伍里最后到达的那个人。在第一种情况下，我们并不知道谁是最后到达的人，因为第ω个人到达的过程会一直持续下去。但是在第二种情况下，我们非常确定谁是最后一个到达的——就是在无穷个人都到达以后出现的那个人。“能够确定最后一个人是谁”的队列和“不能确定最后一个人是谁”的队列是不同的。

在数学上，不能确定队伍最后一个人的情况叫作“极限序数”。因为你已经把一本票据用到了极限，所以你没有办法在不使用另外一本票据的情况下

继续发票据。

注意，存在很多不同长度的队列，其最后一个人都是可以确定的。比如说，所有的有限队列。很快我们就会发现，还有很多不同长度的队列，它们的最后一个人都是不能确定的。

多个无穷队列

在加法之后，我们要尝试的运算是乘法。因为乘法的构建往往开始于重复地做加法。在常规的数字里，我们知道变换乘数和被乘数的位置是没有影响的。所以，

$$5 \times 3 = 3 \times 5$$

更加概括性地说，

$$a \times b = b \times a$$

这个等式对于所有的实数a和实数b都成立。这也被称作乘法交换律。我们已经得到了一些关于该规律对于无穷的数字不成立的线索。那就是在希尔伯特旅馆的问题里，无穷组双人自行车和两辆载有无穷位客人的巴士这两种情况有所不同。

在这个时候，我们需要仔细思考乘法的本质是什么。因为从重复加法的角度构建乘法的方式也有很多种。如果我们的研究对象是常规数字的话，我们是不需要担心这个问题的，因为用不同的方法得到的答案都是一样的。你可能已经不记得存在这么两种思考的方式，这两种思考方式是这样的：5×3到底是“5个3”还是“3个5”？到底是有5个袋子，每个袋子里有3块饼干，还是有3个袋子，每个袋子里有5块饼干？到底是$3+3+3+3+3$还是$5+5+5$？就像之前说的，当研究对象是常规数字的时候，我们不需要担心这些，因为我们知道答案是一样的。然而在帮助小朋友做乘法运算的时候，你还是要花费很多时间才能说服他们同意这一点。这个时候通常的做法就是把一些物体排成下面这样（见图9–2）。

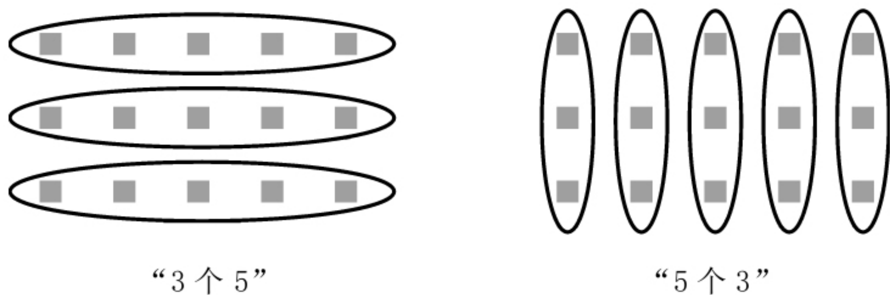


图9–2

因为在无穷序数的情况下情况会有所不同，所以现在我们需要决定“a×b”的定义到底是什么。在数学上通常的定义中，a×b的含义是“a和自己相加b次”。这和ab表示“a和自己相乘b次”在某种程度上是一致的。

所以，对于5×3，我们需要把它想象成5个人到达，又有另外5个人到达，再有另外5个人到达。你需要给出1～15号票据。对于3×5，我们需要把它想象成3个人到达，又有另外3个人到达，又有另外3个人到达，又有3个人到达，再有3个人到达。你还是会给出1～15号票据。

让我们再次思考一下无穷巴士和双人自行车的例子。如果先到了ω个人，然后又到了ω人的话，你就需要用掉一整本红色的票据，然后再用掉一整本蓝色的票据。

相对的，如果无数辆双人自行车到达了，你就要先把1号和2号票据发给骑第一辆双人自行车到达的两位客人，然后把3号和4号票据发给骑第二辆双人自行车到达的两位客人，把5号和6号票据发给骑下一辆双人自行车的客人，以此类推。你可以一直这么做下去，而且不需要使用第二本票据。

在第一种情况下，我们做的是“ω和自己相加2次”，所以我们把这种情况表达为ω×2，它的答案是两整本票据，也就是ω+ω。在第二种情况下，我们做的是“2和自己相加ω次”，根据乘法的定义，也就是2×ω。这一次，答案是一整本票据，也就是ω。

$$\omega\times 2=\omega+\omega$$

$$2\times\omega=\omega$$

因为ω+ω在这里比ω+1要大，所以它肯定已经比ω大了。这也就意味着：

$$2\times\omega\neq\omega\times 2$$

所以，无穷序数的运算并不具备乘法交换律的特性。

你现在可能在想，我们是否能够用上次的方法证明这两者之间的不同。比如看看谁是队列中的最后一个人。然而这次，这种办法是无效的。因为这两个队列都不存在最后一个人。我们需要采用一个稍有不同的方案：我们可以去问每个人排在他们前面的是谁。对于有ω个人的队列来讲，唯一不知道谁排在他们前面的人就是拿到1号票据的人，因为他的前面已经没有其他人了。但是对于有ω＋ω个人的队列来讲，有两个人并不知道谁排在他们前面，分别是：

- 拿到1号红色票据的人，因为他的前面没有其他人了。

- 拿到1号蓝色票据的人，因为所有拿到红色票据的人都在他的前面，但是拿到红色票据的“最后一个人”是不存在的。所以并不能确定排在他前面的人是谁。

通过这种办法，我们能够知道，每次我们增加另外ω个人的时候，我们都会得到一个新的序数。因为每当这个时候，都会有一位新的1号客人，拿着某种新的颜色的票据，且不知道站在他们前面的人是谁。

无穷的减法

我们现在可以开始思考减法了。当我们试图研究无穷是不是一个常规的数字时，我们最初遇到问题的地方就是减法。当我们为整数定义减法的时候，我们是通过加上一个相反数（也就是一个负数）的方式来处理的。然而，序数并不存在相反数。谁是队列中站在–1这个位置上的人呢？你可以想象小孩子们可能会争抢这个位置，觉得这样就能打败站在队伍第1位上的人了。但是实际情况并没有这么简单，除非你能够把这个人添加到另一个队列中并且在那个队列中去掉一个人（也许小孩子们也想这么做。但是到了那个时候，他们说不定已经对整个数学领域感到非常气愤了，就像是你试图让他们通过数点心的数量来学习算术，但是又不许他们吃的时候一样）。

我们需要回到你还是一个小孩子的时候，看看你当初是怎么做减法的。让我们来思考一下5–3。我们都知道答案是2，因为我们能够从5倒着数回去；或者因为我们可以伸出5个手指，然后放下3个手指，再数数剩下的手指；再或者仅仅是因为我们已经对这个答案耳熟能详了。然而，如果你只知道怎么正着数而不知道怎么倒着数的话，你可以从3开始看看数多少数

能够数到5。当孩子们这样做的时候，他们真正回答的问题是“我需要往3上面加多少才能得到5”。我们可以做得更加高级一点儿，把这个问题写成下面的等式：

$$3+x=5$$

还有另外一个相关的问题：“是不是存在这样一个数字，如果我往这个数字上加上3的话我能够得到5？”请留意这两种说法在表达方式上的不同。下面是这个问题的表达式：

$$x+3=5$$

在通常的数字范围内，上面的两个等式是相同的。因为$3+x=x+3$。但是我们已经发现，在无穷序数的情况下，加法交换律是不存在的。所以如果把3或5替换成无穷的话，这两个等式可能就是不同的。在我们思考无穷序数的时候，我们需要分别思考这两种情况。

第一种情况：我需要往1上面加多少才能得到ω？

$$1+x=\omega$$

这个方程可以通过让$x=\omega$来解决，因为我们已经知道：

$$1+\omega=\omega$$

如果我们尝试另外一种情况的话，可能就会出错。是不是存在一个数字，如果我往这个数字上加上1的话，我就能够得到ω？这个问题可以表达为：

$$x+1=\omega$$

这一次我们不能通过让$x=\omega$来解决了，因为我们前面已经知道：

$$\omega+1\neq\omega$$

事实上，这个方程不存在可能的解。因为无论我们把x设定为什么数字，$x+1$都会有“队伍里的最后一个人”。换言之，这个1总是会排在最后。而ω是不会有“队伍里的最后一个人”的。这意味着我们不可能在包含无穷的情况下解出这样的方程：

$$x+a=b$$

这继而意味着我们不能做减法。让我们来看看为什么。

我们需要记住，在无穷的左边加上一些东西和在无穷的右边加上一些东西是不一样的，所以在左边减和在右边减可能也是不一样的。减法就是“撤销”加法。如果我们在数字的左边撤销加法的话，我们就相当于从队列的前面移除一些人。如果我们在数字的右边撤销加法的话，我们就相当于从队列的后面移除一些人。这就是问题所在了，如果队伍里并不存在能够确定的最后一个人的话，我们是没有办法从队伍的最后移除这个人或更多人的——因为我们根本就找不到这个人。这也就是为什么我们解不了下面的方程：

$$x+1=\omega$$

因为为了找到x，我们需要在ω的右边撤销“$+1$”。但是我们可以解另外一个方程：

$$1+x=\omega$$

因为这次我们是在ω的左边撤销“$+1$”。也就是从队列的前面移除一个人。

这就是前面在我们还没有证明无穷不属于任何一种数据类型的时候一直遇到的问题。我们当时说，我们想要：

$$1+\infty=\infty$$

但是那意味着，当我们从等式的两边同时“减去∞”，我们就会得到：

$$1=0$$

现在我们终于找到了一种不会出现这种问题的定义无穷的方法。我们有：

$$1+\omega=\omega$$

我们不能从右边在等式两边同时减去ω，我们只能从左边做减法。如果你想要从有$1+\omega$个人的队列前面减去ω个人的话，队列里就一个人也没有了，这和我们从ω个人的队列前面减去ω人是一样的。所以如果从等式两边同时减去ω的话，等式就变成了：

$$0=0$$

这和我们预想的是一样的。我们终于得到了一个在逻辑上成立的定义无穷的方法，而这个方法也为无穷在减法里的奇怪表现给出了解释。所以我们终于为我的小外甥和他的好朋友之间的争论——无穷到底是不是一个数字——找到了答案。

无穷不是自然数、不是整数、不是有理数，也不是实数。

无穷是一个基数，也是一个序数。

基数和序数并不必须遵守前面的数字类型所遵守的全部规则。这就是为什么在这两个数据类型里，无穷是可能的。

针对孩子们争论的“对无穷倍”的问题，我们也有了一个新的答案。如果他们争论的是序数的话，那么他们所能争论的层次将比基数多得多。在基数的情况下，你必须“以2为底数，以无穷为指数”才能构建下一级无穷。但是在序数的情况下，你只要在无穷上加1就能构建下一级无穷了——你只需要小心地把这个1加在无穷的后面而不是加在无穷的前面。

关于这个结论，我最喜欢的例子来自《驯悍记》（*The Taming of the Shrew*）。

如果这不是你要找的，那么我也无话可说了。去跟比安卡告别吧，比永远还要再多一天。

如果永远是ω的话，永远还要再多一天就是ω＋1。你看，莎士比亚知道ω＋1比ω要更大！至少，我愿意这么去认为。

第二部分　景色

10　抽象事物

我希望我可以瞬间移动。这样，当我想前往一个梦寐以求的地方时，我就能够瞬间到达那里。如果路上风景优美的话，我依然愿意享受一个普通的旅程。但当我仅仅需要到达某个地方、完成某项任务或者拜访一个久违的朋友时，能够瞬间移动将会帮我节省很多时间和精力。

我最喜欢的抽象世界的一个方面就是，你想到的任何事物都能立即发生。比如当我想要探索一个新的抽象世界时，我仅仅需要想到它，瞬间我就身在其中了。再比如，当我想要玩一个新的抽象的玩具时，在我想到它的那个瞬间我就能得到它。只要你有一个想法，这就足够了。因为在抽象世界中，你想到的事物就已经存在了。在抽象世界中，如果我想要我的晚餐出现，我需要做的仅仅是在脑海里打定主意要吃什么。

对于一个抽象事物来讲，“存在”到底意味着什么？对此，我们可以进行长时间的争论。但是对我而言，一个抽象事物“存在”就意味着我能拿它来玩耍。这就是抽象数学研究的工作模式：你想到一个全新的数学理论观点，然后你立即就可以开始在脑中与它游戏。围绕着这个概念组建新事物，观察它和其他事物之间的相互作用。也许它会引起矛盾，使一切崩溃。但是这个观点依旧存在，你仍然能继续和它玩耍。这不像你有了一个关于新型汽车或者新型药品的想法，在那种情况下，你需要完成一系列的工作来将它转变成“现实”，包括筹备资金来购买设备和材料，反复试验，等等。

举个例子，数字的观念之所以存在就是因为我们已经思考过它了。那么，这是否意味着数字真的存在？这取决于你问谁。哲学家对此有一个长期的争论。就我个人而言，我更乐意认为数字仅仅是一个观念。我对数字是否存在的担心并不比我对自己是否存在的担心更多。无论我是否“存在”，我仍将继续我的人生。无论数字是否“存在”，我仍将继续开展数学研究。我对于存在的态度可能确实有点儿与众不同，从下面的例子中也可见一斑：我是我认识的所有人中唯一相信圣诞老人存在的人，因为圣诞老人是人们在圣诞节收到礼物的基础。一些人也许会说这仅仅意味着圣诞老人这个概念是存在的。但是我认为，说圣诞老人是一个抽象的观念，这是一种褒义的说法，就像数字是一个抽象的观念一样，这不影响我对它的喜爱。

在这个意义上，我很乐于说无穷是存在的——以一个抽象的观念的形式存在。但是我们还是想知道，无穷的存在形式能不能不这么抽象。我们是否能在“真实生活”中找到一个无穷量级的东西？出于几个原因，我经常不那么愿意使用“真实生活”这个词组。首先，抽象事物并不必然比“真实”事物缺乏真实感。比如，疲惫是抽象的，但对于我来说，疲惫的感受是非常真

实的；而太平洋的海底世界是真实的，但它对我来说则是非常抽象的，因为我无法触摸、观察或者体验它。其次，所谓“真实生活”中的数学问题常常是完全不合乎常理的，比如你的野马逃跑了，或者出去买75个西瓜什么的。

宇宙中是否存在一种无穷量级的“真实物体”，这是值得怀疑的。当然，极大数量的分子、原子或者电子是肯定存在的。但是我们已经分辨过了，“极大数量”和“无穷”之间存在着巨大的差异。对于我们小而有限的大脑而言，宇宙本身看起来似乎就是无穷的，但它实际上很可能是有限的（当然这个问题的答案目前还是未知的）。

可以肯定的是，存在无穷数量的抽象对象。以数字为例。我们知道自然数是绝对没有边界的。从这个方面讲，这个概念和我们无法确定其是否有边界的宇宙这个概念不同。每一个独立的自然数都是有限的，但是它们能够不断通过加1来增大直至永远。从数字的存在这方面而言，它们的数量是无穷的。在下一章，我们将看看另外一些类似的事物，它们也能够持续增大以至于我们无法为它们设置边界。像自然数一样，它们在任何给定的瞬间都是有限的，但是它们能够增长到如此之大，以至于我们有必要把它们看作“接近无穷”，并且，用这种方式来理解它们自有其道理。我有一个朋友收养了一只大丹犬幼崽，它长大的速度快到令人难以置信，以至于在某一刻它看起来似乎要接近无穷了。

在第12章和第13章中，我们将会看到，虽然我们的大脑本身不是无穷的，但是它们可以拥有无穷的潜力。我们分辨微妙事物的能力是没有限制的，在第12章中，我们将凭借这种能力抽离出更高维度的空间。这些高维度的空间不同于物理上的空间，因为它们存在于思想领域而非物理领域。我们将看到，我们可以对于一本书的故事情节给出“稍稍有点儿一维”这样的非正式的评价，而这种评价是有意义的。也就是说，如果我们在这个层面上思考维度，那么我们的生活将是非常多维化的。维度的数量不可能真正被限定，所以我们将又一次接近无穷。在第13章中，我们将会讨论一种不同类型的维度。这种维度类型来源于范畴论，这是一个抽象数学的领域，它研究不同事物之间的关系。在这里，我们思考事物之间的关系，接着研究那些关系之间的关系，再之后是那些新关系之间的关系，持续下去直到“永远”，这将引导我们进入无穷维度类别。所有这些维度类型都会为我们提供更加微妙、更加有表现力的途径来帮助思考我们周围的世界。

接下来，我们将换一种方式来看待无穷。比起一些非常大以至于没有边界的事物，我们可能更想要了解一些事物的微小部分。我们意识到，如果我们思考事物的无穷小的部分，我们会发现所有事物都可以分解成无穷多个无穷小的部分。在第14章中，我们将看到对于无穷小的事物的思考是如何导致了一些奇怪悖论的产生，使得人们花费了数千年的时间才把它们解

决。在第15章中，我们将发现，在解决这些难题的过程中，数学家们意识到他们可能并不真正了解什么是实数。至少他们的理解没有准确到足以让他们形成能够站得住脚的论断。在第16章，我们将观察到一些奇怪的事物，它们来源于我们对无穷和无穷小的事物的新理解。最后，在第17章，我们将研究无穷是以什么方式含蓄地出现在我们的日常生活中的。比如，通过建立循环，我们就能够永远环绕着前进。但是，“永远”到底是什么意思呢？

永远的抽象版本

在日常生活中，当我们谈论到某个事物能够“持续到永远”的时候，无穷就出现了。我们有“持续到永远”的小数和“持续到永远”的自然数。在真正的现实生活中，没有事物能够持续到永远。但是在数学里，我们有一个抽象的理解角度。这个角度能够帮助我们立即获得永远。这是抽象带给我的众多乐趣之一。

在这本书的开头，我提到过我最喜欢的计算机小程序。这个小程序能够在屏幕上输出无穷行“你好”。它只包含两行代码：一行是启动程序，另一行告诉计算机“再做一遍”。这就是我们抽象地获得永远的方式。这个游戏和我们之前构建自然数的方式有相通之处。我们从1开始，然后持续对其加1直到“永远”。我们是抽象地在做这件事，而不是真的重复加1这个动作。如果我们通过用实际行动来加1，然后再加下一个1，然后再加下一个1，这将需要我们花费“永远”那么长的时间，而这个时间比我们的生命长得多。然而，如果我们只是抽象地来做，那么我们就能马上得到所有的自然数。这就像在原理上做而不是在实践中做一样。这就是抽象世界的有趣之处，无论是在原理上做还是在实践中做，我们得到的结果都是一样的。两种方式真的没有多大区别，不像那些令人哭笑不得的说法，“理论上讲，理论和实践是相同的，但是在实践中……”

我们用不断加1的方法构建所有自然数的方式就是数学归纳原理的基础。从本质上说，它告诉你的是原则上你可以同时考虑所有的自然数，即考虑数字1以及考虑在所有数上加1。

这里实际上是在说，如果你想表明某些事物对于每一个自然数都是正确的，你不必针对每个自然数去证明它，你只需要证明：

- 一个起始点，也就是对于数字1，它是正确的；

- 一种向前推导的方式，也就是如果对数字n它是正确的，那么对于数字

$n+1$它也是正确的。

一旦你完成了上面两步，那么你一下子就会明白这个事物对于所有自然数来说都是正确的。

就像一个小孩子，一旦他发现自己已经学会如何向上爬一级台阶，他需要的就只是有某个成人将他随意放在楼梯的底部，然后他就能沿着楼梯向上爬了。他可以想爬多高就爬多高，直到他被大人抱走。原则上，他是可以永远向上爬的。但是在实际中，即使他不被抱走，他也会感到饿和疲惫。数学对象并不会感到饿和疲惫。当我们处在数学情境中做某件事时，一旦原则上我们可以多做一遍，这就意味着我们可以做这件事无穷次。这就是在数学中某些事物能持续到“永远”的意思。数字能够永远持续下去，数列能够永远持续下去，小数也能够永远持续下去。

我们拥有无穷数量的自然数，这个事实也帮助我们从两个方向上接触到其他各种类型的抽象的无穷事物：更大的无穷和更小的无穷。我们已经看到了我们可以从无穷的自然数集开始，构建逐渐增长、越来越大的无穷集合。但是在我们生活于其中的世界里，我们在哪里才能找到这样的无穷事物的集合呢？一种方式是考虑无穷小的事物。这是微积分的基础，它是应用范围最广的数学领域之一。微积分是对处于变化中的事物的研究。对于总是在变化的事物进行理论化研究是很困难的，而微积分是通过考察无穷小的部分，再将无穷多的这些无穷小的部分黏合在一起来完成这项工作的。这些无穷小的东西在我们周围随处可见。事实上，我们每天都在无意识的情况下接触由无穷小的东西组成的无穷集合。在这本书的第二部分，我们将看一看无穷在我们周围出现的方式，无论我们是不是已经意识到了它们的存在。

数学有一个奇怪的特征，就是无论我们是否看到了它或者理解了它，数学都在那里。就像我们坐在火车上，无论我们是否往窗外看或者是否知道刚刚经过的是什么地方，风景都在那里。了解数学能够帮助人们构建更有效的工作体系，并解决更复杂的问题，同时，它还能揭示我们是如何与周围世界相互作用的，并且阐明我们的思想是如何运作的。与解决一个具体的问题或者开创一个具体的技术相比，启发是数学带给我们的一个更为微妙而又不那么引人注目的好处。对我而言，启发那默默无闻的重要性意义深远。

11　从千层酥到iPod

我几乎可以确定，我是不会去爬珠穆朗玛峰的，虽然我会乐观地保留一点儿能瞬间移动到那里的可能性。我也几乎可以确定，我不会去南极。我不认识任何一个攀登过珠穆朗玛峰的人，但我确实认识一位在南极工作的天体物理学家。我知道到达南极是非常困难的，即使是坐飞机也很困难，但南极和我们之间的距离仍然是有限的。我也知道珠穆朗玛峰的高度是有限的。但是，它们对我来说就像无穷远或无穷高一样，因为我永远也不会去那些地方。

无穷是存在的，但是我们能到达那里吗？如果要做的工作无穷小的话，我们是否能做无穷多的工作？在我们搞清楚这些问题之前，让我们先考虑一些看起来非常大，似乎要接近无穷的事物，以及一些看起来我们几乎是在做着无穷的工作的情况。

有一个关于在国际象棋棋盘上放米粒的古老难题。这个难题的背景是这样的：一个人被要求在国际象棋棋盘的第一个方格里放上一粒米，在第二个方格里放上数量翻一番的米，在第三个方格里，米粒数量再翻一番，以此类推，直到棋盘上的每一个方格都被放满，一旦他完成这项工作，他将得到整张棋盘上的米。问题是，最终他能得到多少粒米？简化版的答案是，非常多。但是准确地说，到底一共有多少粒米呢？

原则上讲，这不是一个困难的问题。你只要不断乘以2，然后将所有得到的答案加在一起，直到你计算完所有的64个方格就可以了。然而，如果你尝试这么做，你会发现数字增大的速度快得惊人。在正常情况下，这个数字会比你的计算器甚至普通模式下的计算机所能处理的最大的数字还要大。除非你有一些特殊的计算工具，否则你是无法算出这个数字的。有一些窍门可以帮助你加快计算，但是你最终还是需要处理一个非常非常大的数字——18446744073709551615粒米。

当然，我们通常不会以粒来计量米，除非在听起来有点儿荒谬的数学问题里。（我第一次在数学课堂里听到这个问题时，曾经试图通过手工计算来得出答案。事实证明，我算错了。）那么上面那个米粒的数量在实际情况中相当于多少米呢？我们仅需要称量出1克重的米，然后数出粒数——1克米大概有50粒左右——然后我们就可以近似地进行下面的计算：

1 克		= 50 粒米
1 碗	= 100 克	= 5000 粒米
1 人	= 每天 4 碗米	= 20000 粒米
全世界	= 70 亿人	= 140000000000000 粒米
1 年	= 大约 500 天	= 70000000000000000 粒米

我们最后得到的数字有16个零。我们曾经计算出的米粒数量为18446744073709551615，近似于在2后边有19个零。这个数字比我们算出来的够全世界所有人吃一年的米多出来3个零，换言之，大约是其1000倍。所以这么看起来，这些米可以养活全世界人口大约1000年。（当然在这里我们没有考虑到的真实情况是，世界人口每年都会增长很多。）

我的计算是非常粗略的，但是我相信你能明白我的意思：只不过是在一个棋盘上移动，做着一些平淡无奇的数量翻倍，但是很快，你就拥有了无法计量的米粒——比世界上现有的所有的米还多。

千层酥

千层酥的制作依赖相同的原理，即重复做乘法使事物以极快的速度增长。千层酥里具有惊人数量的薄层，它们是通过将面团折三折这个步骤重复6次制成的。第一步，把一层厚厚的黄油夹在面团中间。当黄油的用量刚刚好时，黄油可以在你擀面团时恰好平铺在夹层里。第二步，将面饼折三折，然后冷却。这样所有层都会保持一定的硬度，而不会和其他层粘在一起。第三步，再一次擀面，将面饼再次折三折，然后冷却。如此操作6次。重复“×3”使得层数增长得非常快。接下来，当你烘烤千层面团时，薄薄的黄油层融化了，黄油的液体部分蒸发并产生蒸汽，这将层与层相互推离。因此除了抽象数字的增长以外，你还能看到千层面团在烤盘中有形地变大。

这是我最喜欢的关于指数增长的实物演示。一般情况下，人们说事物正在呈指数型增长，仅仅是想表达它们增长了很多。这基本上是正确的，但其正式的数学定义是事物一直以相同比例的速率增长。如果我在第一次将面团折三折，第二次折四折，第三次折五折，第四次折六折，这样层数会增

长得更快，但它就不是指数型增长了，因为增长的速率发生了变化。

我喜欢指数增长能够直接对应于美味的千层酥这个事实。千层酥的多层次不仅仅是为了引人注意和美观。当薄薄的一层层酥皮在口中融化的时候，它能够带来精妙的口感。千层酥以难以制作闻名，但是我认为使用指数增长的天才方法重复折三折的步骤，实际上大大降低了制作那些千层酥薄层的难度。毕竟，单独擀出这样的薄层是非常困难的。数学的主旨就是将复杂的事物变得简单一些。不幸的是，不知怎么，它反而往往会创造出莫名其妙的复杂问题。

iPod

我记得当iPod（苹果公司发布的便携式数字多媒体播放器）刚刚问世的时候，我在地铁上看见了它大大的广告标语——“240首歌曲，100万种不同的方式”。这个广告的核心理念是希望让人们记住一件事，那就是就算只有240首歌曲，如果你不是在播放列表中从头播到尾而是用随机的顺序来播放它们，那么你将有100万种不同的播放方式。

事实上，100万是一个极大的低估。出于好玩的目的，我坐在火车上计算了一下为了获得100万种不同的播放方式，你最少需要多少首歌。

如果你有两种歌，那么你只有两种播放它们的方式：从一首开始然后播放另一首，或者反过来。假设你有三首歌。对于第一首歌，你有三首歌可供选择；对于第二首歌，你只有剩下的两首歌可供选择；对于第三首歌你将不再有选择。（这里有一个前提假设是一首歌你只想听一次，虽然在现实中我经常重复播放同一首歌曲几个小时。）我们要再画一张树状图。与之前不同的是，这一次每一层的枝杈会越来越少，因为你不想重复播放，所以我们可供选择的歌曲数量就会越来越少（见图11–1）。

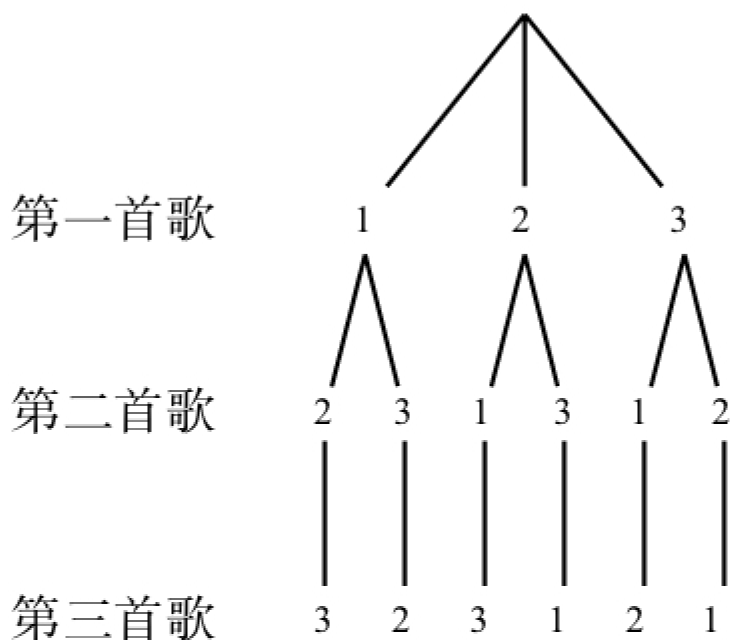


图11–1

所以在这里我们有6种可能的播放顺序，从树的最上层到最下层的每一条路径都代表了一种播放顺序。根据每一级可供选择的路径的数量，我们可以把可能的播放顺序总数计算为$3 \times 2 \times 1$。

现在假设我们有4首歌，我们会得到图11–2这个树状图：

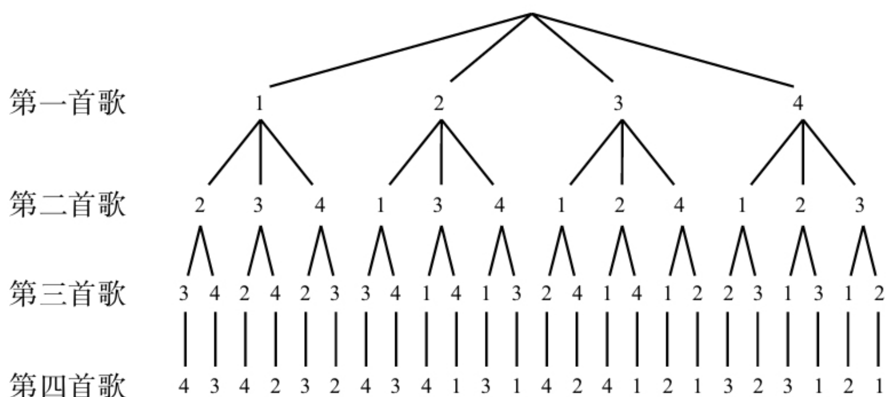


图11–2

或者通过下面的方式计算：

- 第一首歌有4种可能性，
- 第二首歌有3种可能性，
- 第三首歌有2种可能性，
- 最后一首歌有1种可能性。

我们可以计算出，总的可能性的数量为$4\times3\times2\times1$。（事实上我们并不需要在最后写上1，因为乘以1并不会改变任何事。）

这在数学中被称为阶乘。“4的阶乘”写为“4！”。一般来说，n的阶乘为：

$$n\times(n-1)\times(n-2)\times\ldots\times4\times3\times2\times1$$

我们也可以用归纳法来定义这个概念。这看起来有点儿像我那个“你好”小程序里面的循环：

- $1!=1$
- $(n+1)!=(n+1)\times n!$

所以，如果我们有n首歌，可能的播放顺序总数就是n！。因为第一首歌我们有n个选项，第二首歌有n–1个选项，第三首歌有n–2个选项，以此类推，倒数第二首歌有2个选项，最后一首歌只有1个选项。

现在，我们可以回到最初的那个问题：为了获得100万种不同的播放方式，你至少需要多少首歌？以数学的方式来说，这意味着我们要寻找一个最小的n值，并使这个n！大于100万。我们可以依次写出阶乘的值，直到我们找到这个数字。记住，从一行到下一行计算下去的时候，我们只需要用这一行的结果乘以下一个n的值：

$$1! = 1$$

$$2! = 2$$

$$3! = 3 \times 2 = 6$$

$$4! = 4 \times 6 = 24$$

$$5! = 5 \times 24 = 120$$

$$6! = 6 \times 120 = 720$$

$$7! = 7 \times 720 = 5040$$

$$8! = 8 \times 5040 = 40320$$

$$9! = 9 \times 40320 = 362880$$

$$10! = 10 \times 362880 = 3628880$$

好极了，我们已经突破了100万！我们只需要10首歌曲，就能用超过300万种顺序播放它们了。

我们现在可以问另一个问题了。如果有240首歌，我们能有多少种播放它们的方式？我们需要做如下的计算：

$$240 \times 239 \times 238 \times \ldots \times 3 \times 2 \times 1$$

如果不使用特殊工具，这个计算几乎是不可能实现的。因为这个数字会变得非常非常大。简单地使用普通计算机电子表格尝试了一下之后，我发现在采取近似之前它只能计算到17！。于是我们得到：

$$17! = 355687428096000$$

我高兴地发现我的手机计算器能计算得更多：

$$18! = 6402373705728000$$

我的手机计算器可以用近似的方法计算接下来的阶乘，但是它最多只能告诉我：

$$103! \approx 9.9 \times 10^{163}$$

之后就完全没指望了。如果再进行更大数据的计算，它只会提示我出现错误。如果采用近似计算的话，我的计算机电子表格能一直算到170！：

$$170! \approx 7.3 \times 10^{306}$$

接下来它也没指望了。这是因为我的旧而脆弱的计算机是使用内存来储存大的数据的，而171！是它碰到的第一个因为太大而无法处理的阶乘。

计算统计学家里克·维克林（Rick Wicklin）编写了一个计算大数阶乘的程序。他最近恰好在他的博客里公布了200！的数值。他解决大数字存储问题的方式是让计算机把大的数字当作一个长长的字符串来处理。然后他让计算机像人类手算一样进行乘法计算，一位一位地做并且在需要的时候进位。他计算200！得到的答案就是下面这串超乎寻常的数字：

78865786736479050355236321393218506229513597768717326

32947425332443594499634033429203042840119846239041772

12138919638830257642790242637105061926624952829931113

46285727076331723739698894392244562145166424025403329

18641312274282948532775242424075739032403212574055795

68660226031904170324062351700858796178922222789623703

89737472000000000000000000000000000000000000000000000

00000000000

我相信这是我见过的最大的数字。它有375位数字，比广告中描述的“100万种方式”要大非常、非常、非常多个数量级。而事实上，这还不是240！。

在摆弄这些大到荒谬的数字时，我还想看看，你的iPod里需要有多少首歌才能让它们的播放方式多于那个棋盘上的米粒。答案是，21。这就告诉我们，翻倍数字能使它们变大得非常快，而阶乘能让它们变大得更快。

你也许想知道在那个数字的末尾有那么多0是否是一个奇怪的巧合。事实上，上面那个数一共有49个0。大数的阶乘结果注定会在末尾有大量的0，即使不知道阶乘的答案，我们也能算出它的末尾会有多少个0。最终答案中的每一个0都来自一个因子10，而每一个10都来自阶乘数字中的一个因子2和一个因子5。由于存在的因子2要比因子5多，因此我们只需要数出从1到我们所计算阶乘的数字之间有多少个5的倍数。记住，有些数字具有多个因子5，因此会贡献多个0。

你长得有多快？

小时候，孩子们会长得非常快。因为他们实在是太小了，所以看起来会比实际长得更快一些。假设一个孩子在最初的几年里一年大约长高10厘米，对于一个婴儿来说，这个增长量显得非常大，因为10厘米占婴儿总长的比例很大。当然，孩子们的生长速度是不同的。有些孩子在小时候比较矮，但是未来某段时期会出现一个急剧的增长，并开始超越周围的所有人。

在数学中，我们也会考虑事物增长得多快，是否有些事物会比其他事物增长得更快这些问题。例如，在米粒的例子里，我们用2^n来表示数量，这里n是稳定增大的。在歌曲播放的事例中，我们用n！来表示数量，这里n是稳定增大的。在这两种情况下，数字都会非常快地增长到难以置信的大小，虽然它们仍然是有限的。在数学上，我们说“当n趋于无穷时，2^n趋于无穷”。在这里，我们非常谨慎地避免了直接宣称某些事物“是”无穷的。同样，当n趋于无穷时，阶乘n！也趋于无穷。但是我们会感觉到n！比2^n增长得“更快”一些。这又是什么意思呢？

搞清楚这件事的一个方式就是把二者构建成一个分数：

$$\overline{2^n}$$

从这个分数中，我们可以观察随着n的增加谁会“赢”得比赛。如果分数数值持续变大，就意味着n！胜利了。如果分数数值持续变小，那么2^n就胜利了。如果分数保持不变，就表示这是个平局。现在我们可以再次耍一个小聪明，将这个分数写成：

$$\frac{n \times (n - 1) \times (n - 2) \times (n -}{2 \times 2 \times 2 \times 2 \times \cdots \times 2 \times 2 \times 2 \times 2}$$

接下来，我们可以像这样将它分解成一个个单独的分数相乘：

$$\frac{n}{2} \times \frac{n}{2} \times \frac{n}{2} \times \frac{n}{2} \times \cdots \times \frac{4}{2} \times \frac{3}{2} \times \frac{2}{2} \times \frac{1}{2}$$

现在，我们可以看出这是一大堆分数的乘积，它们几乎都是头重脚轻的。

除了最末尾的$\frac{1}{2}$。此外，随着n越来越大，分数的乘积逐渐包含越来越多的

项，同时新项将会比之前的项更为头重脚轻，因为其顶部的分子将越来越大，而底部的分母则始终是2。最终的结果就是顶部明显胜过底部。

事物趋于无穷的速度是有一个等级的。这与我们之前所说的无穷的层级有一些不同，但都是基于相似的想法建立的。

我们刚刚看到，n！增长得比2^n快，而2^n又比n^2增长得快。事实上，2^n比n^3、n^4或n的任何次幂，甚至$n^{10000000000000000000}$增长得都要快。最后一个数看起来很大，而且除了n＝1以外，在开始阶段它确实比2^n要大。但是2^n最终会超越它。我们来试一个稍微不那么夸张的数，比如n^{100}。我的计算机告诉我，n^{100}会一直领先2^n直到n＝125。之后，2^n就领先了。

在这里，我们只考虑n的正数次幂。因为负数次幂通常根本不会带来增长。随着n的增大，n的负数次幂通常越来越小。

有个东西比所有可能的n的幂都要增长得慢，那就是log n。你也许还记得，对数是与指数相反的。如果我们以10为底来做对数，那么log n就是“将10增大到n所需要的指数”。所以log 100就是2，因为你计算10的2次幂就能得到100。而log 1000就是3。然后100到1000之间任意数的对数都位于2和3之间。以10为底数取对数基本上就是在计算一个数在以10为基数的情况下包含多少位数字。所以当n增大时，log n也会持续增大，但其增大的速度很慢。当n已经增大到100万时，log n才仅仅增大到6。

这就是对数有用的一个原因。它可以将大数字转换成小数字，这样我们就能较方便地处理这些数字了。有这样一个理论，当数字增大到超过某个特定点的时候，我们的大脑就无法真正处理这个数字了。所以我们转而使用

对数的方式思考，观察这个数有多少位而不是看它有多大。你可能也会这么做，即使你没有意识到你是在“取对数”。这就是为什么在之前的例子里我会告诉你200！所得到的数字有375位，因为375是一个我们能处理的数字。换言之，我刚刚从对数的角度对这个数字进行了转换。当一个数字大到这个程度时，与数字的数量级相比，你基本上已经不会在乎在这个数的某一位上加上或者减去一个1了。这就是为什么在米粒的例子中我如此随意地假设一年中有500天。我知道相对于这个事件的宏伟性，500与365之间并没有什么重要的区别，我真正需要做的就是选择一个与365具有相同位数的数字。这里，我就使用对数的方式进行了思考。

总而言之，对数比n的任意确定的幂都要增长得慢。

慢速增长

有可能你增长得非常缓慢，以至于虽然你还在增长但是已经看不出来了。

想象一下你正在节食，你决定只吃一块蛋糕的$\frac{1}{2}$。但是蛋糕太美味了，你决定再吃这块蛋糕的$\frac{1}{3}$。接下来，由于你像我一样控制不住地想要更多但是又真的不想变胖，所以你决定再多吃一点点，但是这次只吃这块蛋糕的$\frac{1}{4}$。接下来$\frac{1}{5}$，再接下来$\frac{1}{6}$，一直继续下去。最终你将吃下多少蛋糕呢？经过一段时间之后，你每次吃的蛋糕将基本上不存在了。因为在100万轮之后，你只能再吃这块蛋糕的百万分之一，这基本上就等于什么都没吃，对吧？

错了。如果你持续这样吃下去直到永远，你最终将会吃下无穷量的蛋糕。事实上，你吃的蛋糕量是呈对数增长的。它的增长速度确实很慢，而且会越来越慢，但仍会不可阻挡地趋于无穷（见图11–3）。

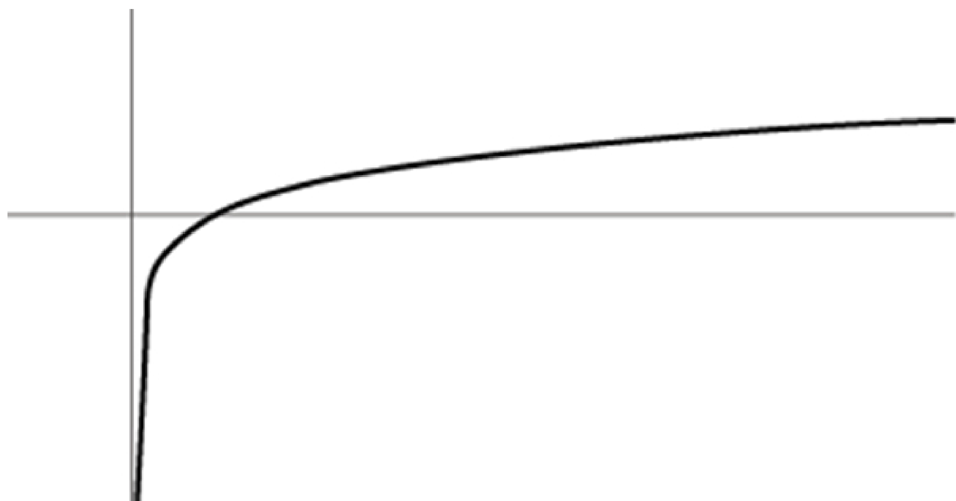


图11–3

从图11–3中你可能无法判断这个曲线最终是否会完全变平，但我来告诉你，它真的不会。这个曲线是无边界的，这表示无论你想到哪个数字，最终它都会超过那个数字。无论你想要它变成多大，最终它都会变得比你预想得更大。我们将在第16章解释这个问题。

这个例子说明，我们在思考无穷的增长的时候，真的要非常小心。因为我们的直觉极有可能导致我们误入歧途。一块蛋糕是微不足道的，但如果我们想要吃到的是无穷量的蛋糕呢？这就是我们需要依赖数学的精密性来搞清楚这些明显令人困惑的矛盾，并且确定我们能否吃到无穷量蛋糕的原因。

12　维度问题

想要来一场时光旅行吗？这听起来很令人兴奋，但是同时也很恐怖，因为你只要稍稍干扰了你自己的过去，就会引起可怕的结果。这些循环和潜在的矛盾激起了我的好奇心，而且是我最喜欢的一些虚构作品的主题。比如电影《回到未来》、小说《时间旅行者的妻子》，以及近期的电影《环形使者》。这些作品的剧情对我来说太过“烧脑”，以至于在看电影的同时我不得不在维基百科上查阅简介。

如果你只想进行简单的时光旅行并且在旅行期间不与其他任何人互动的话，那么这个旅行潜在的危险后果就小了很多。这可能听起来毫无意义，但是如果你正在被坏人追逐，这个小知识会非常有助于你摆脱他们，比如你可以逃去第四维度。在这一章，我们将会思考世界上到底存在多少个维度。

想象一下你在火车上，而有人正在追捕你。为了困住你的火车，他们需要做的就是在火车前面和后面设置障碍。火车是卡在轨道上的，所以它不可能绕过障碍物——它只能在一个维度上移动，而不是在两个维度上。

然而，如果你在一辆车里，你就可以开车绕过障碍物，所以坏人为了抓住你需要在你的车周围设置一整圈障碍物。这时候，如果你想逃离，你可能就需要变成詹姆斯·邦德，按下一个按钮将你的车变成一架飞机，这样就你能飞离包围圈了。换言之，你可以逃往第三个维度。

现在为了抓住你，他们需要在空中围绕你布下一整张网。这一次，你需要使用第四维度才能摆脱他们……

我们已经习惯于生活在一个三维世界中，所以很难想象维度更多的世界是什么样的。你也许会觉得，“并没有更多维度了”。如果我们只考虑物理维度，这句话是正确的。但那只是思考维度的一种方式——一种非常实际的、物理的方式。实际的例子对于掌握一种新观点是非常有用的，但也有其局限性。就像通过数饼干来理解数字很简明，但是我们很难用这种方法思考负数，因为很难想象“负数的饼干”是什么。

思考四维的一种方式就是把它作为三维的归纳。为了更好地理解这种归纳的运作方法，我们可以退几步，从起点开始思考。如果我们能观察清楚我们是怎样从一维走到二维，接下来又是怎样从二维走到三维的，我们就能研究怎样从三维走到四维、从四维走到五维，然后从任意一个n维走到n＋1维。就像孩子爬楼梯，或者我们沿着无穷这个梯子前进一样。这样一

来，也许只要我们不停下，我们就能到达无穷数量的维度。（有时候我觉得这种归纳到永远的思路在某种意义上是一种数学乐观主义。）

如果你生活在一个一维世界里，那意味着你基本上是生活在一条直线上。当然，这并不一定意味着你生活在一条物理的直线上，它只表示你只能在一个方向上移动。向前走意味着你朝着这个方向的正向前进，而向后走意味着你朝着这个方向的负向前进。另一个思考方式是，你可以通过发送一个坐标来告诉某人你的位置。无论一条路是笔直的还是弯曲的，沿路的房屋都能够按顺序来标记。即便是在一条圆形道路上，你仍然只需要告诉别人一个坐标就能让他们找到你：你只需告诉他们你沿着这条圆形道路距离他们有多远。你也可以给他们两个坐标，告诉他们你处在相对他们向北多远向东多远的地方，但是这种做法既没有效率也没有必要。例如，你可以说你在沿着圆环距离公认起点逆时针45°的位置，而不是说你在向北50米且向东50米的位置上（见图12–1）。

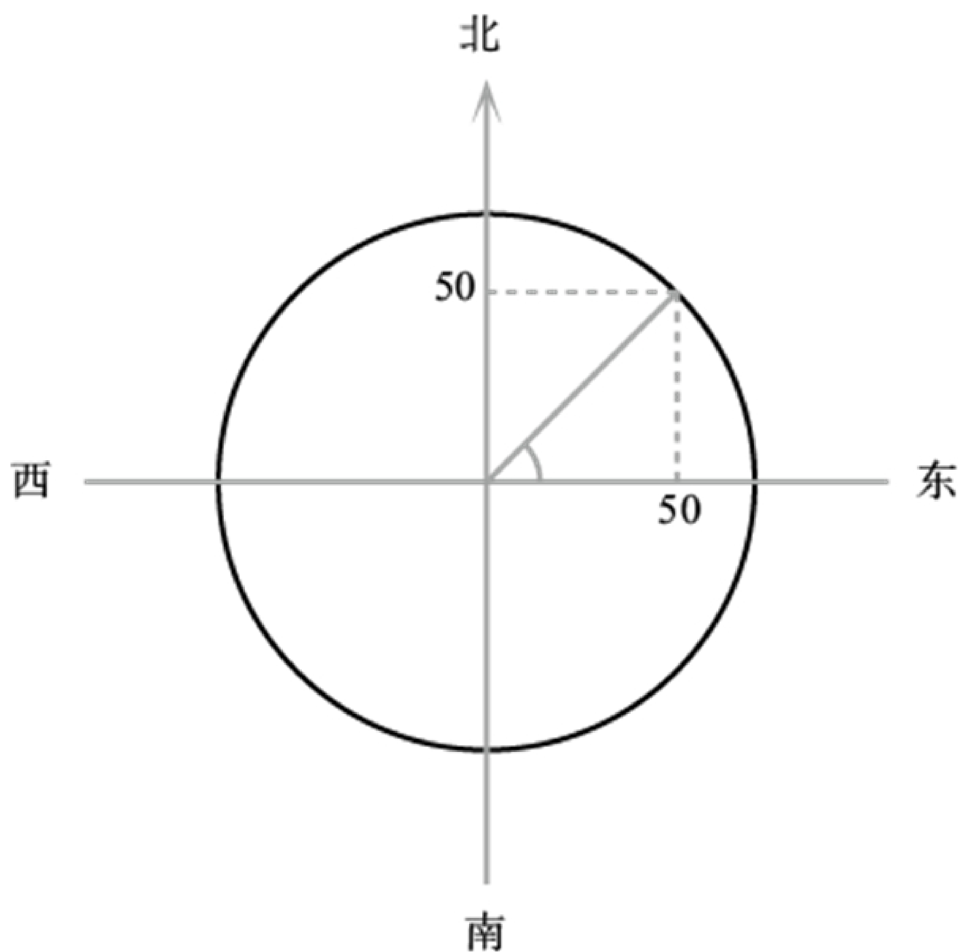


图12–1

一维世界就是那么直截了当。这里只有一个变量，就是距离多远。然而令人吃惊的是，在一维世界中我曾经多次迷路。我在尼斯大学工作时，我所在的数学系系楼是环形的，但不知怎么我总是找不到我的办公室。当我试图从卫生间回到办公室时，我常常会迷路。虽然抽象地讲，向前和向后都是在同一维度上，但是在实际生活中，这二者是截然不同的。在我试图找到我的办公室时，思考数学是毫无帮助的。

构建一个二维世界能帮助你从一维世界中逃离。离开火车（或者像詹姆斯·邦德一样从窗户跳出来）就是这种情况。一维世界的问题在谢菲尔德这个城市表现得非常充分。在那里，一维的电车与二维的汽车共享街道。当然，电车和汽车自身都是三维的，我在这里指的是它们移动的维度数。如

果一辆汽车在电车轨道上抛锚了，其他的汽车只需要绕过它就可以了，但是电车在坏在它所行驶的路线上的汽车被拖离之前，是完全无法移动的。它甚至不能向后移动，因为那样它的行驶路线就错了。在一维世界里困住某些东西是非常容易的，你只需要一个障碍物就行了。这就是为什么人们过去常常在城堡周围建护城河，因为这样去往城堡的路就基本上只剩下那座一维的吊桥了，而一座吊桥是非常容易切断和守护的。如果你没有护城河，你就需要在城堡四周都进行全方位的防御。

你很容易就能分辨出来你现在是否身处二维世界，因为在二维世界中，你需要两个坐标来表明你的位置。你需要明确指出你“向前方”行进了多远，以及“向侧方”行进了多远。这就是为什么电影院的座位和飞机的座位都有一个排数以及一个座位数。从原理上讲，因为座位数是有限的，你也可以从1开始往上给它们设编码。但是这样做的话，寻找座位会变得比较困难。

地球表面是个非常有趣的事物，因为它看起来是三维的，但实际上，它是二维的，因为你只需要两个坐标就能说明你在哪里，这两个坐标就是经度和纬度。当然，这里没有考虑你身处地下的情况。球体表面是二维的，虽然它只能存在于三维的宇宙空间中。这是一个提示：“维度”并不像它看起来的那么直接。哪怕我们讨论的只是物理上的维度。这就好像虽然圆圈是一维的，但是你仍需要一张二维的纸才能画它一样。一条翻山的小路也是一维的。但是它在指向上和下的同时还会向左、向右扭动，所以它也只能在三维的空间里存在。二维空间中的两个维度通常使用笛卡儿坐标系来表示，即包括x坐标和y坐标。这和芝加哥的笛卡儿网格系统是一样的（见图12–2）。

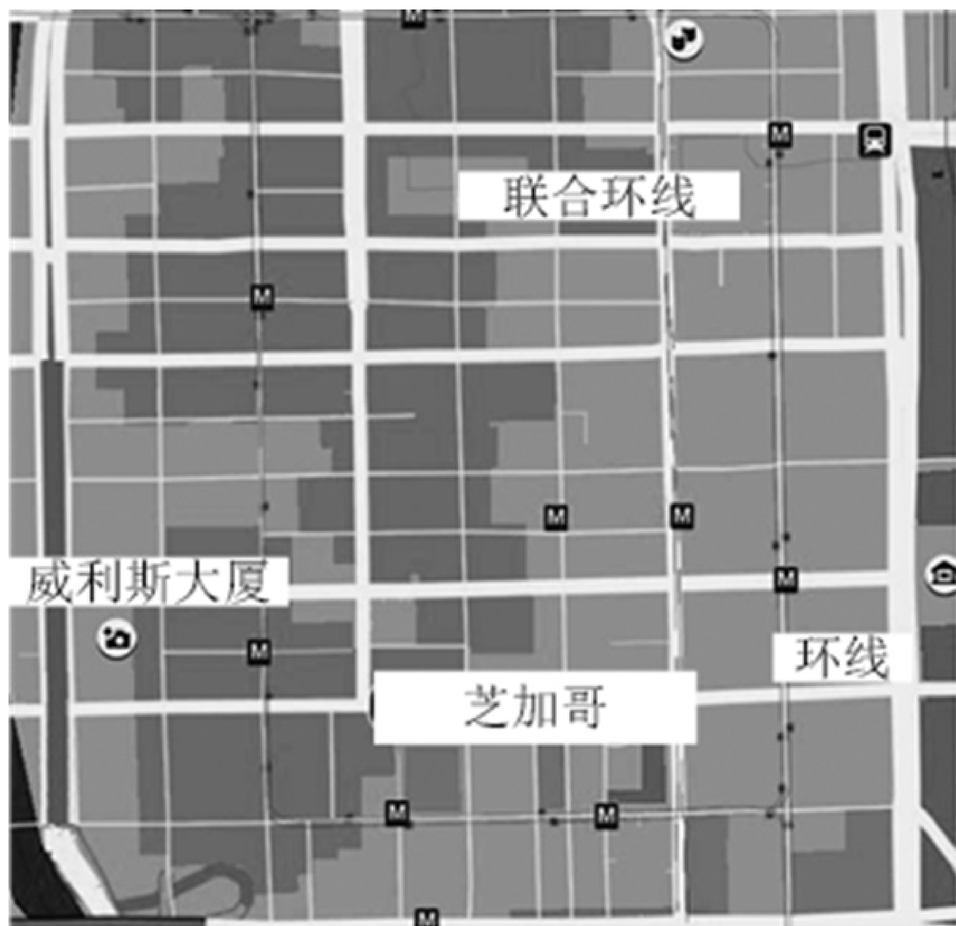


图12–2

然而，我最近在阿姆斯特丹居住过一段时间。那里的同心运河体系使城市形成了一个极坐标网格（见图12–3）。在这里，你的位置不再用东西南北来表示，而是用你所处的角度和你距中心的距离来表示。

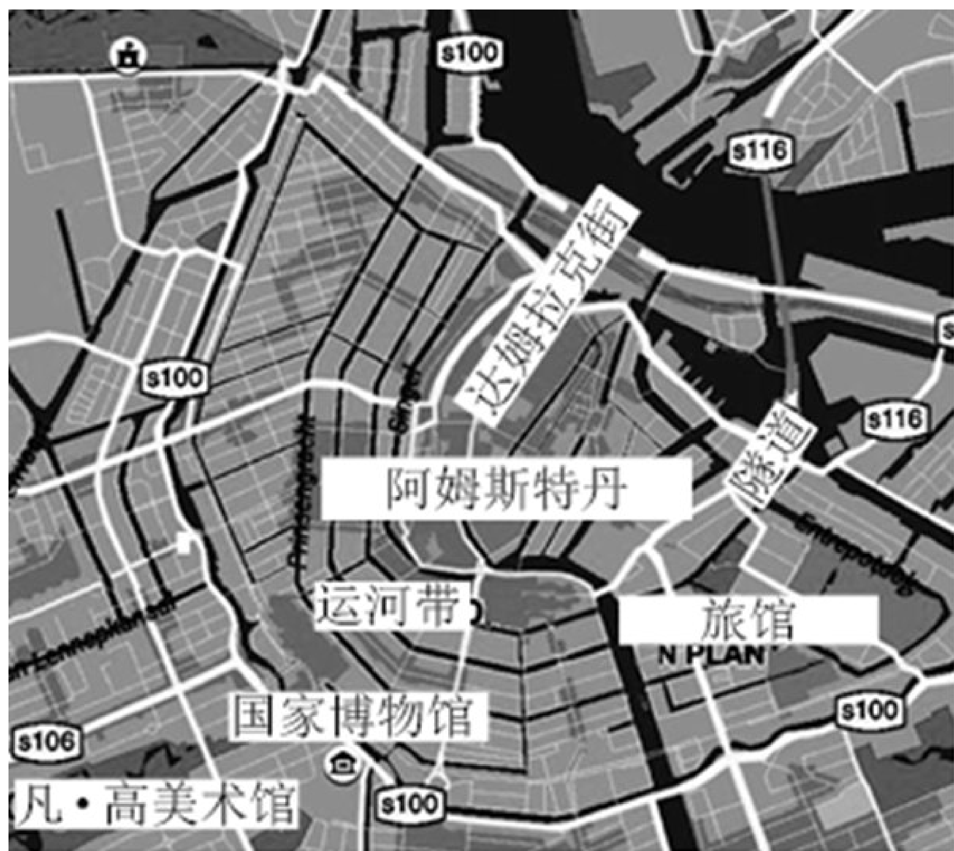


图12–3

三维世界让你能够逃离二维世界。我喜欢在水下游泳的一个原因就是我能够体验三维的自由。我猜想这也是滑翔和跳伞的乐趣之一，虽然我从来没有尝试过这两种运动。同样我也无法太过自由地感受第三维度，因为我受到重力的约束。三维的存在解释了为什么你只需要一圈栅栏来困住一头牛，却需要一整个笼子来困住一只鸟。这也是飞机战斗比汽车追逐要更为复杂的原因。一架飞机需要三个坐标来确定它的位置，经度、纬度，还有它正在飞行的高度。同样，这就是为什么你用一个套索就可以抓住一头牛——好吧，也许你不能，但是有些人能——但是你需要用一张网来抓一条鱼或者一个超人。（当然，对付超人，你还可以用一些氪石。）

四维世界能够帮助你逃离三维世界。如果一条鱼能打开一条神奇的通道进入第四维度，它就能逃离你的网。如果你是詹姆斯·邦德，而你在飞机上被坏人包围了，你也可以按一个按钮逃向第四维度。

可能的第四维度

如果你觉得这很难理解，你可以把时间想象成第四维度。这是思考时间的一个十分有效的方式。这种方式在理论物理中得到了充分的运用。（但是这只是思考第四维度的其中一种方式，而不是唯一方式。）这意味着，如果你被抓了，你可以通过时间旅行来逃离。这是时间旅行类小说中最常见的主题。在《回到未来》中，马丁意外地通过时间旅行的方式，从射杀博士的人手中逃离。在《时间旅行者的妻子》中，亨利被逮捕了，但他也意外地依靠时间旅行逃跑了。

思考时间的另一个方式就是把它当成另一个你需要明确的坐标。如果你想和某人约会见面，你在确定地点的同时还要确定时间。如果你们都去同样的坐标，但是是在不同的时间去的，你们也无法相遇。你们将会处于相同的地点，但是你们在“空间—时间”这个坐标里并不处于同一位置上。

下面我将展示你可以如何使用时间作为第四维度来逃脱追捕。我们先来思考你如何使用第三维度从第二维度中逃离。如果你被围墙围住，逃离并不难，方法就是翻过围墙。用围墙困住你的人错误地认为你只能在两个维度里移动，也就是说，他们认为你不能更改你的垂直坐标。所以你需要做的就是简单地改变你的垂直坐标，翻过围墙，然后你的垂直坐标会再次变回零。现在，你的横向和纵向坐标就都在围墙外了。换言之，你现在安全了。

如果某个人用网抓住你，你将需要更改你的时间坐标来逃脱。你只需要简单地通过时间旅行到达昨天就可以了，因为昨天你还未被抓住。之后，你可以稍稍改变一下你的空间坐标，然后通过时间旅行回到今天。现在你就安全了，因为你的物理坐标已经在网外面了。

另一个思考四维的方式就是把第四维度想象成颜色。也就是说，只有当你与某物具有相同的横向、纵向和垂直方向的坐标以及具有相同的颜色时，你与该物才真实地处于同一地点。你可以通过将自己涂成不同的颜色来改变自己的颜色坐标。这意味着某个人若想要用一个白墙的房间困住你，他们就必须将你涂为白色。为了逃离，你只需要获取其他颜色的染料，比如紫色，然后把自己涂成那个颜色，接下来，你就可以直接穿过白墙了。当你安全地穿过墙壁后，你就可以转换成任意你想要的颜色了。这有点儿像隐身衣。不同之处在于隐身衣只有两种状态——可见和不可见，而且即使你是不可见的，你也不可能穿过墙壁。

我最近将上面这些关于四维的解释说给一位音乐家听，他立即说，“这是不是说，我们也可以把音乐看作第四个维度？因为音乐可以帮助你逃离三个物理维度。”我说，是的。虽然这有一些复杂，因为音乐并不能帮助

你“永久”地逃离。如果你被困在一个房间中，你可以通过听音乐逃离被困的感觉，但是你并不能使用它来使自己物理地离开那个房间。如果我在火车上听着我的音乐，而另外某个人听着一些完全不同的音乐，这就像我们完全不在同一个地点一样。如果我刚刚欣赏了一场气势恢宏的音乐会，在回家的路上，我会感觉我和没有听过这场演唱会的人处在完全不同的维度里。这也像时间旅行一样，因为当我听到一段音乐的时候，我立即就能感觉到自己仿佛身处另一个与那段音乐紧密相连的时空之中。

你也许想知道为什么维度必须与逃离相关。原因是如果你不逃离原先的维度，你就不可能得到一个真正的新维度。事实上，如果你愿意，你总是可以用更多的坐标来描述你自己的位置。你可以说你在向东多远、向北多远、向东北多远的地方。比如你可以说，“我在向东4英里、向北2英里、向东北$\sqrt{18}$英里的地方”（见图12–4）。

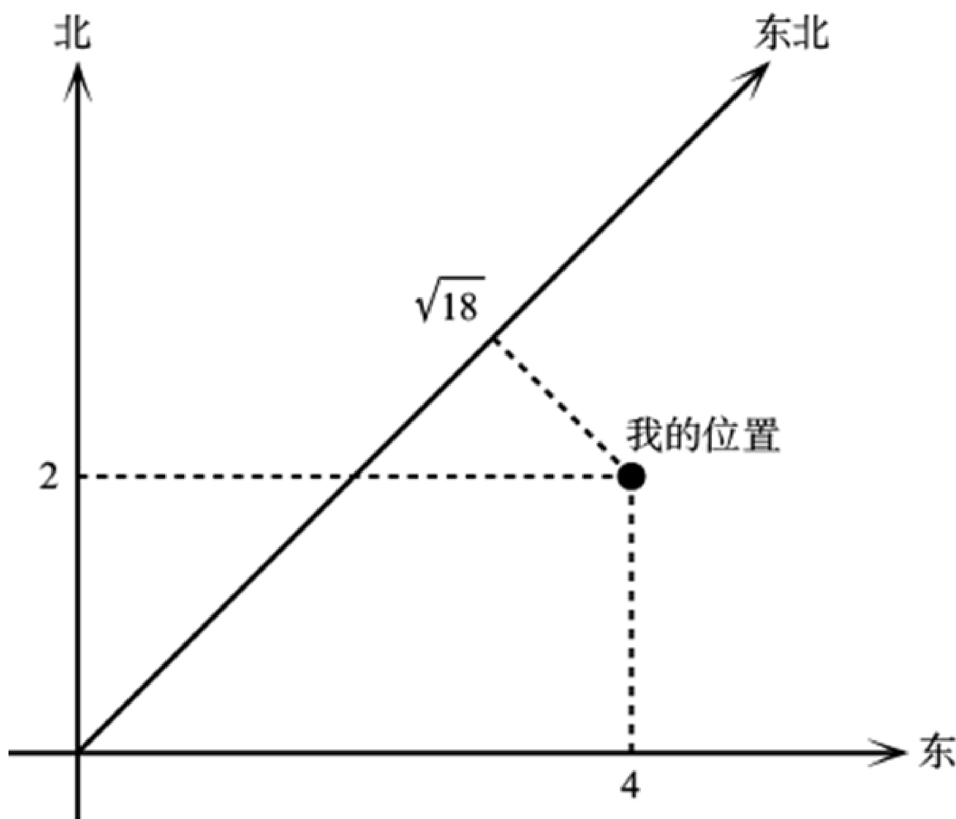


图12–4

这个说法非常准确，但是非常多余。在数学里，这涉及维度“独立性”这个概念。如果其中一个维度能够通过其他的维度来表示，它就不是独立的。在我们的例子中，东北这个坐标就不具有独立性。（这在语言学上是显而易见的，因为“东北”就是一个由“东”和“北”组成的词。）

真的有具有四个维度的事物吗？

因为我的研究方向就是高维范畴理论，所以我花费了很多时间正式或非正式地与人们讨论维度。当我给非专业人士演讲时，听众中往往会有人觉得这个内容非常令人心烦意乱，并且坚持认为我的演讲是毫无意义的。因为他们根本不相信存在第四维度，因为我们就生活在一个三维世界里。

从物理上讲，他们的说法是正确的。但是一旦你开始根据独立坐标来思考维度，那么事情就完全不同了。我在前面介绍过，当我们从数学的角度来想象某个事物，那么这个事物就存在了。所以，只要我们能想象出多于三个的相互独立的观念，那么我们就能创造出一个高于三维的思想空间。我确信你在生活中想出过超过三个的独立观念。就像莫里哀在《贵人迷》中揭示的，一个人终其一生都在散文中诉说却没有意识到这一点。同样，你可能在你大部分的人生中都在做多于三个维度的思考，只是你并没有意识到。

归根结底，坐标只是一串数字。数字能够表达任何可以衡量的事物。距离就是这样一个可衡量的事物。如果使用更加字面化的表达方式，我们可以说，重量（或者质量）是一个你能够在某个范围内测量的事物。上次我去看医生的时候，他们记录了我的年龄、身高、体重、脉搏和血压。鉴于血压包含两个数字，那这里就一共有六个数字了。如果你想绘制一个包含每个人的体检信息的示意图，你就需要做一个六维的图。

我最近花了几天时间练习制作法式马卡龙。这非常复杂，因为其中包含了太多变量。虽然它需要的原料种类很少，但你还是需要决定每100克蛋清中需要使用多少糖霜、细砂糖和杏仁粉。接下来，你需要决定搅打蛋清要用多长时间、翻拌面糊要用多长时间、要挤出多大尺寸的圆形、花多长时间把它们晾干、烤箱设置成多少度以及烘烤它们要多长时间。这是一个九维空间。而我甚至还没有把购买鸡蛋时的复杂情况考虑进去，因为购买鸡蛋这个过程中也有它自己的变量，比如尺寸、颜色、是不是来自散养鸡、是不是有机的、是不是适合素食者的。我希望挑选大的、棕色的、来自散养鸡的、有机的、适合素食者的，但是我常常到结账时才发现我漏掉了检查某个或者某几个变量。

每一次你根据一个标准列表来比较事物的时候，你事实上就是在观察一个具有许多维度的空间。从这个角度来说，只有三个维度这么少的空间是很

少见的。一旦你感到有必要列出一个标准表，你面对的维度就很可能超过三个了。下面这个例子就很常见，即使你并没有注意到你自己列出了这个单子：每次买机票的时候，我都会权衡价格、航空公司、航班时刻表、中转次数以及机场，但这些我都是在头脑中完成的。

机械臂

你可能会说，那又怎么样？真的有必要去研究那些更高维度的空间吗？甚至，我们真的需要知道它们的存在吗？虽然你一直没有察觉你正在思考更高维的空间，但你还是生活得很好。就像是散文的主人公虽然不知道自己身在一篇散文中，但不也一样诉说了那么多年吗？

研究更高维空间的一种方式就是利用时空这个概念。时空将我们平常的三维空间与作为第四维度的时间组合在一起。爱因斯坦相对论中的一个观点就是时间是可以弯曲的。

比时空更实际可见的是对机械臂的研究。机械臂遍布各个领域，包括工厂、外太空、微创手术和街机游戏。机械臂本身是在三维空间里移动的，但是为了研究它的移动范围，你需要考虑在任何一个既定时刻它的每个铰链或者关节的动作。每一个铰链都是一个变量，所以你最终得到的空间的维度数会与铰链的数量相同。如果你的机械臂用的不是铰链而是更加复杂的关节的话，你用到的维度数量还会更多。

花一点儿时间观察一下你自己的手臂。如果你坐着挥动你的手，你的手看起来就是在三维空间中移动的。你是怎么做到的呢？这同样依赖于各种铰链和关节做着不同类型的运动。

- 如果你保持从你的上臂到你的手之间各部位相对静止，然后挥舞你的手臂，你会发现你的肩关节本身会为你提供两个维度的运动。换言之，为了说明上臂相对于身体的位置，你需要两个坐标，也就是上下和前后。

- 你的肘关节提供了一个维度或者坐标，也就是你的前臂与你的上臂之间的角度。

- 你的手腕提供了两个维度。也就是你需要两个坐标来确定你的手相对于前臂的位置：一个坐标是上下，另一个是左右。

- 还有前臂的旋转，也可以说成手的位置：手能够从掌心向上旋转为掌心

向下。

另外还有你上臂的旋转。你可以将上臂保持在相对于你身体的同一位置，并固定你前臂与上臂间的角度，在这种情况下继续旋转。你在向别人招手时或多或少会用到这个姿势，这种类型的招手是用手在空中划出弧线，而不是拍手。

这里一共有七个维度。但是我的手真的是在七维空间中移动的吗？

如果你穿过裙子，你可能有过要花费很长时间将背后的拉链拉上的经历。如果你从来没有穿过裙子，你也可能在试图往后背涂抹防晒霜时遇到过相似的状况。在拉拉链时常常会发生的情况就是，你可以用手从身后将拉链从底部拉起，但是到某个特定位置你就会卡住。接下来你将不得不改变方式，将你的手绕过肩膀，然后从上方将拉链拉到顶部。如果你的柔韧性不够，那么这件事你就无法完成，因而必须去寻求帮助。往后背涂抹防晒霜也是相似的情形。你可以用手从身后涂抹下半个背部，但是到达某个点，可能是肩胛骨附近，你就不得不停下来，然后你需要将你的手绕过肩膀，从上方继续涂抹你的肩膀和上半个背部。同样，如果你的柔韧性不够，在背部中间就会出现一块没涂到的区域，那里很可能会被晒伤。

这里发生的情况就是，存在一个你的手在较低位置所能到达的边界。如果你有足够的柔韧性，那么这个边界将会刚好与你的手从上边伸到背后所能到达的边界重合，甚至可能重叠。在这两种情况下，你的手在三维空间中都到达了相同的地点。但是你的手臂使用的是完全不同的配置，并且在这两种配置间转换需要花费很多力气。这意味着在手臂配置的七维空间中这两个地点相隔非常远。这类情况就是我们在设计机械臂时需要考虑到的。我们确实需要知道“手”在普通三维空间中能到达的地方，但是更为重要的是，我们需要考虑在手臂移动配置的高维空间中哪些位置是紧密相邻的。假设你在微创手术中需要微微移动某个组织一段很小的距离，但是机械臂为此需要进行彻底的重新配置，那将是很不合适的。

降低维度

考虑高维空间是一件困难的事。这就是为什么数学中有一整个大的领域致力于研究它。这也是为什么在生活中和在数学中我们常常会努力减小我们考虑的维度数，以此降低问题的难度。做到这一点有很多种方法。

一种方法就是直接忽略一些维度。举个例子，如果你正在根据标准列表评估某些内容，你可能会意识到你面前的标准太多了，你应该将重点放在那些重要的标准上。此时，忽略一些最不重要的标准能够瞬间降低你思考的

维度数，让你更容易地做出决定。

你也可以通过固定某些变量来暂时忽略一些维度。当我想要优化我的马卡龙时，我最终决定每次只在一个维度上研究它们。我做了一批相同原料配比的马卡龙，然后在不同的温度下烘烤它们。接下来，我固定温度，然后制作了一批以在混合原材料的不同阶段挤出的混合物为原料的马卡龙。再后来我固定这两个参数，然后稍稍改变杏仁粉的用量。这就像在普通的二维空间里进行操作，并且将y坐标固定在某个特定的值上，比如2。然后你剩下的就是这条直线了。在这条直线上，x可以是任何数，但y必须是2（见图12–5）。

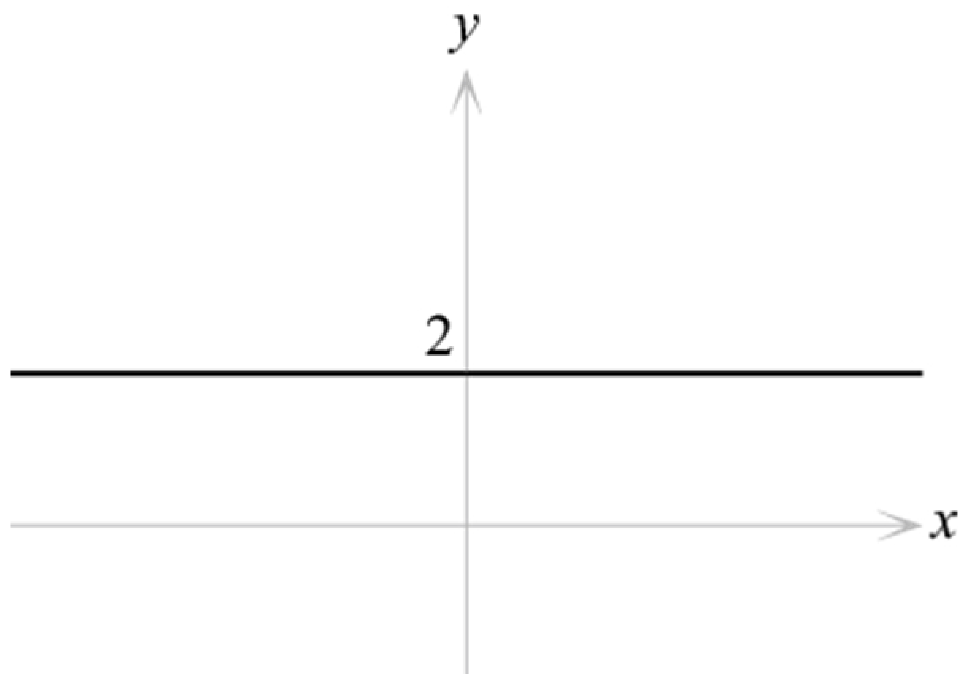


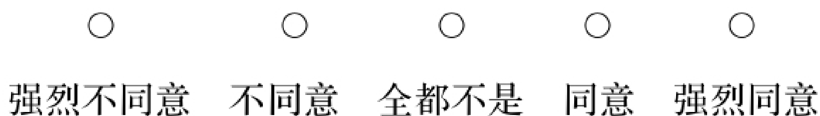
图12–5

通过这种方法，我们将二维的平面转化成了一维的线。这与完全忘记一个变量没有太大不同。如果你彻底忘记了一个变量，就等于将问题降到了x轴，也就是把y固定为0。

在制作马卡龙的案例中，实际过程会更复杂一点儿。因为我并不知道一个配方的最佳温度是否会是另一个配方的最佳温度。所以在改变了杏仁粉用量后，我还需要再一次尝试所有不同的温度。你可以想象，这项研究将会

花费很长的时间。

我们经常会用一种更为微妙（也更为狡猾）的方法来降低维度，那就是将两个维度合并为一个维度而不是仅仅忽略一个维度。这在编写调查问卷中时常发生。当调查者问你是否同意类似于“比起一对一社交我更喜欢大群体社交”这样的观点时，你被要求接下来必须按照下面这样的标准来表达你同意或是不同意：



理论上这一调查结果应该比直接让你回答是或者否更加贴近真实答案，因为事物不再只有黑色或者白色，我们允许这之间存在灰色地带。然而，这类问题常常让我沮丧，因为我经常发现自己相同程度的非常同意和非常不同意。从表面看，我应该选择“全都不是”，但是这感觉也不对，因为“全都不是”听起来似乎表示我并不真正在意，而实际上有些日子我是真的想要融入一个大团体中，而另一些日子我是真的想与人一对一地交流。因为这本质上是个二维的问题：你有多喜欢在大团体中社交，以及你有多喜欢一对一社交？这里包含两个变量，我们可以据此画出类似图12–6的示意图。

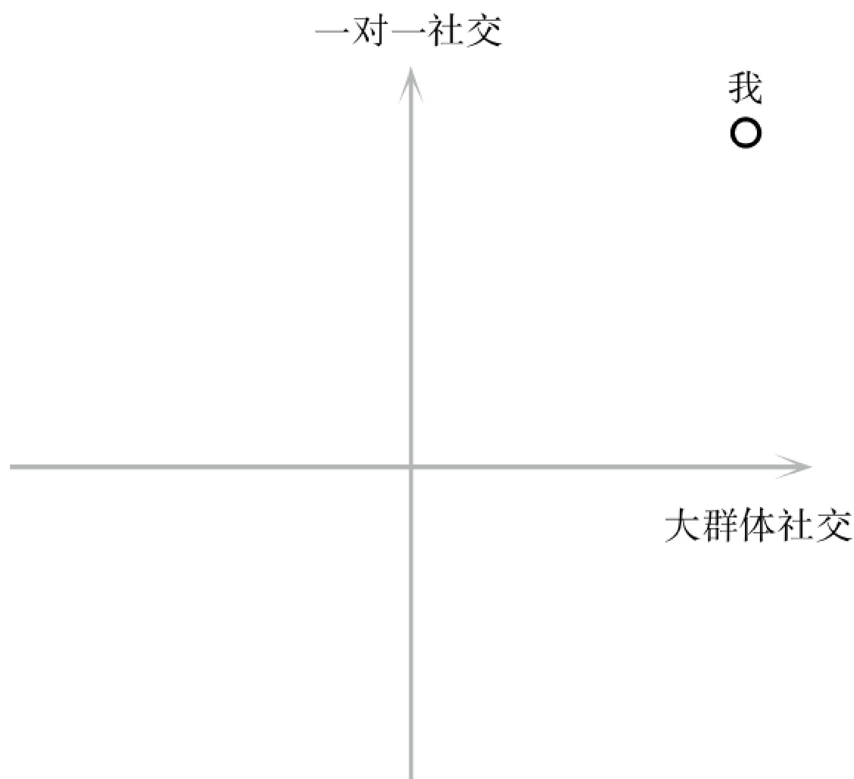


图12–6

因为我两种社交方式都很喜欢，所以我就位于这个坐标轴中右上角的某个地方。但有些人两种社交方式都很讨厌，那么他就位于左下角的某个地方。事实上，整条对角线表达的都是相同程度地喜欢和不喜欢这两种社交方式（见图12–7）：

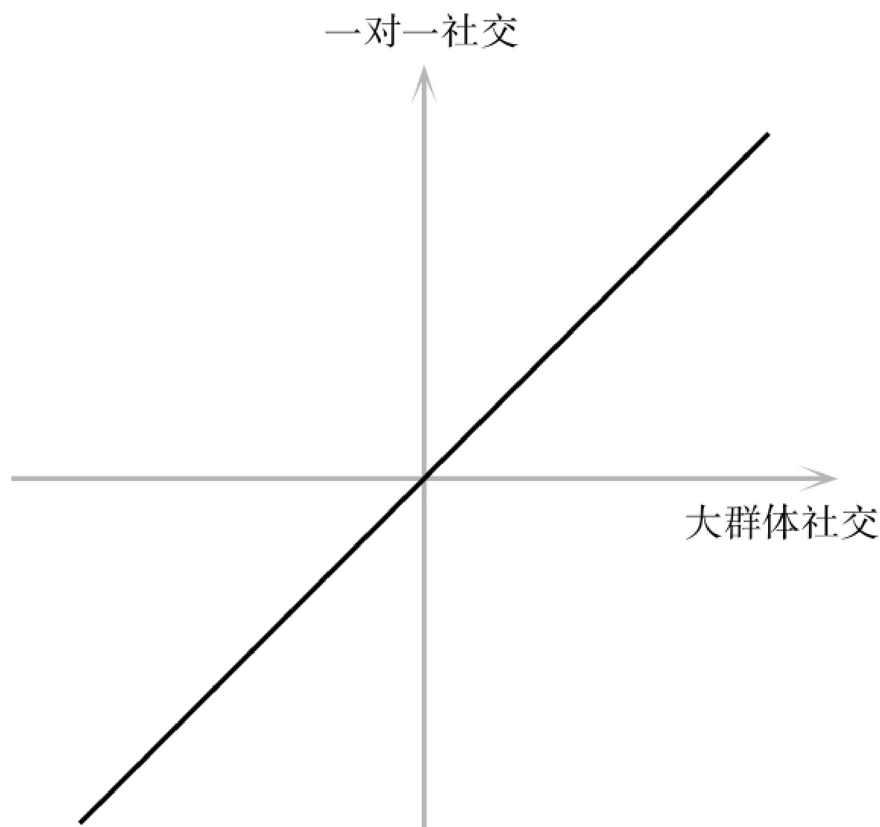


图12–7

这条线在数学上的表达式是$x=y$。在原来的问卷设计中，这条线被压缩成一个简单的点——“全都不是”。这个问卷所做的就是假定大群体社交和一对一社交是“完全相对立”的事件，我告诉他们我喜欢其中一个的程度就等于自动告诉他们我不喜欢另一个的程度。这就像一个零和游戏。零和游戏在数学上就表示为$x+y=0$，也就是图12–8这条线：

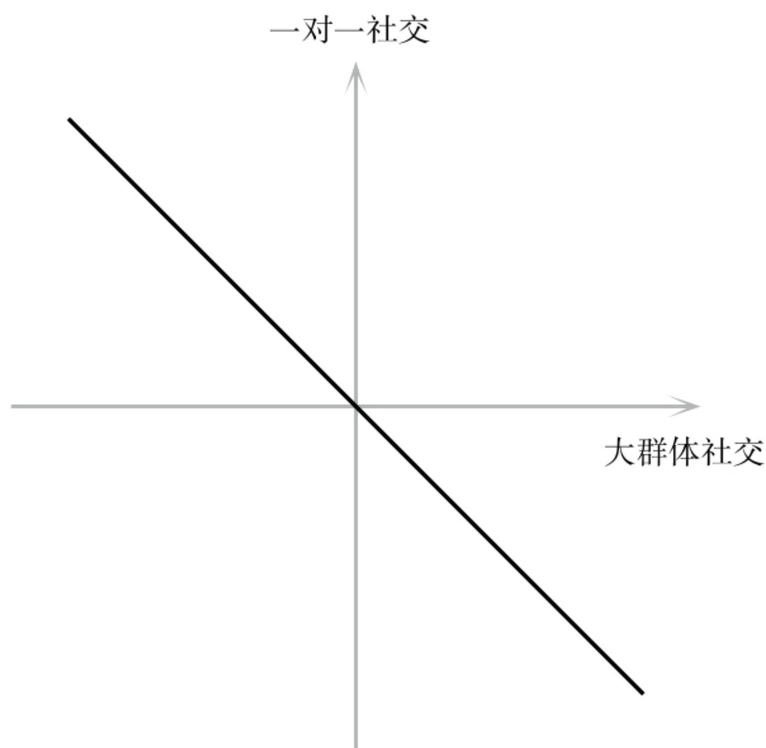


图12–8

我们在这里遇到的情况更像是一个“加和为4的游戏”，因为总和总是为4。为了达到4，我认定你对于某个观点可以在0～4这个范围内表示同意或者不同意。这样一来，这个问卷调查看起来就是在假定，如果你认同一对一社交的程度为n，则你认同大群体社交的程度就是4–n，就像下面这样：

一对一社交	4	3	2	1	0
大群体社交	0	1	2	3	4
	○	○	○	○	○
	强烈不同意	不同意	全都不是	同意	强烈同意

如果我们将这些点画在图中，就会得到图12–9：

图12–9

有时候，调查者会将问卷做得更加细腻，承认人们的感受是连续的。在这种情况下，他们允许你在一条从“同意”到“不同意”的线上的任何一个位置做标记（见图12–10）。

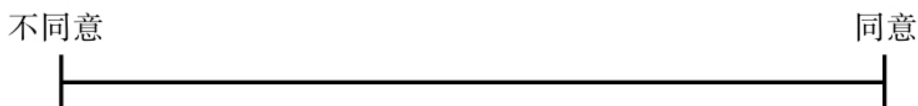


图12–10

在这种情况下，我们面对的就是x＋y＝4这一整条线，表达为图12–11：

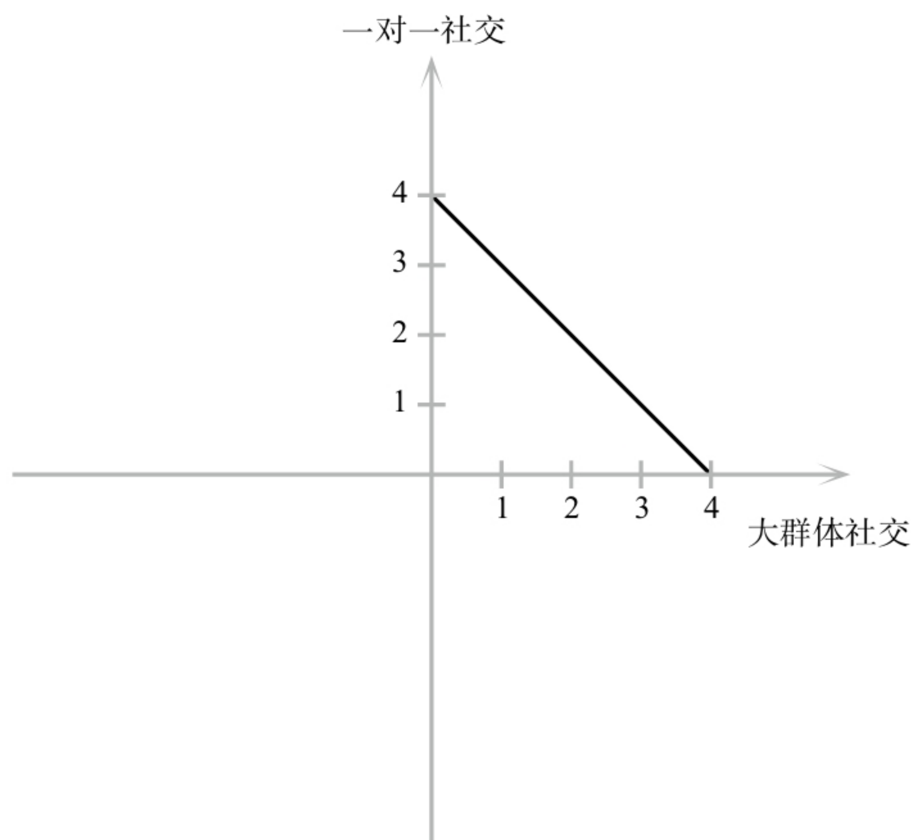


图12–11

你可以在这条线上任意选择一点，如果你得到了它的x坐标和y坐标，你会发现它们加起来是4。如此我们就用了一个稍微有些奇怪的方式将两个维度压缩成一条一维的线。

另外一种在调查问卷中降低维度的方式就是将我们喜欢一种社交方式的程度与我们喜欢另一种社交方式的程度相比。在这种情况下，我们将x坐标和y坐标简化为$\frac{x}{y}$。这是将两个维度降低为一个维度的一种更加复杂的途径。你可以找一个制作三维示意图的在线工具，然后试着用这个工具做出

$z = \frac{x}{y}$，从而观察每个点在二维平面中对应的位置。这是很难直观想象的。

降低维度也经常发生在政治信仰的语境中，我们倾向于谈论人们有多“左翼”或者多“右翼”。这是一种过度简化的方式。检测政治倾向的网站politicalcompass.org指出，确实存在两个直接变量，经济观点和社会观点，通过这两个变量，我们可以绘制出一个二维的政治倾向图。但是在实际使用的过程中，人们总是把这个二维图压缩成一个一维的句子。简化既有优点又有缺点。优点是简单的事物更容易掌握，缺点则是你会因此而丢失信息。我们常常需要权衡简化的优点和缺点，或者至少我们应该意识到我们正在做简化这件事。

事实上，政治信仰常常处在更高维度的空间中，因为当我们在描述自己的政治信仰时，我们往往不会对变量数量进行限制，甚至会让自己陷入一个有无穷维度的空间里。但是我们不得不减少维度，这样我们才能尝试将自己编入一个有大致相同的观点的群体中。困难出现在我们试图通过合并维度而非忽略它们来降低维度的时刻。比起决定哪个特定标准是不重要的，我们更常选择假定某些标准之间是相关的。从数学的角度讲，这两种减少维度的方法是非常不同的。

维度的连续性

有时候，当我对某种情况进行“利弊”评估时，我会发现标准与标准太难区分开了。例如，你可能会考虑短期效益和长期效益。接下来你开始想知道中期效益。我还会考虑微期效益，也就是瞬间期许。那么短期是在哪里转化成了中期？这些标准的边界在哪里？

可能你正在评估不同的工作，你会考虑工作满意度还有自我价值实现程度。但是这两项标准是有重叠的，因为你在工作中的自我实现也会提升你的工作满意度。也许，自我价值实现程度只是工作满意度的一个方面？于是你开始将工作满意度分解成越来越多的更详细的标准。但是当你这么做的时候，它们之间的边界也会变得越来越模糊。

最终你会发现，你制定的标准本身也是一个连续体。这就意味着，你不仅是在连续体中评估每一个标准，而且标准本身也处于一个连续体中——你实际上处在一个维度数量不可数的多维度空间中。难怪做决定是那么困难！

13 范畴论和结合体

我的一个朋友在社交媒体上发布了一条关于“美味牛角包三明治”的信息，我立刻就把它想象成了一个夹着牛角包的三明治。但很快我就意识到，她说的牛角包三明治是把牛角包当作三明治的外层而不是夹心。

你是否曾想过制作一个“三明治三明治”？也就是制作一个夹心就是三明治本身的三明治。也许在叠起来的面包切片之间应该放些生菜。有些人坚持认为生菜对于三明治来说至关重要，它发挥着结构性的作用，并且可以保护面包不被填料浸湿。（虽然我自己从来不会主动选择吃生菜。）在这种情况下，x三明治的组成应该像下面一样：

面包

生菜

x（夹心）

生菜

面包

例如，一个鸡肉三明治应该是这样的：

面包

生菜

鸡肉

生菜

面包

在这种情况下，鸡肉三明治三明治就应该是：

面包

生菜

鸡肉三明治

生菜

面包

如果我们把它完全展开，就会变成：

面包

生菜

面包

生菜

鸡肉

生菜

面包

生菜

面包

这有点儿像我最喜欢的一种蛋糕——“迭代巴腾堡蛋糕”。你可以先做一个巴腾堡蛋糕，它的样子就像下面这样（见图13–1）：



图13–1

然后，你可以做一个每个小方块都是一个巴腾堡蛋糕的大巴腾堡蛋糕（见图13–2）：



图13–2

也许在大蛋糕里的小巴腾堡蛋糕周围也应该有一圈杏仁蛋白膏？

这两个例子本质上都是用事物本身来搭建事物。数学特别擅长用其自身来建构事物，比如高维空间就是由低维空间组建而成的。这是因为数学能够处理像空间、维度和无穷这样的抽象概念，而且它本身也是一个抽象的概念。物理对象则不会有这种表现。如果你将大量的鸟放在一起，你并不会得到新的鸟，这根本不是鸟类生成新物种的方式。那么事物抽象到什么程度才能够迭代呢？

用乐高拼乐高

想象一下用小乐高块来制作大乐高块。这是可行的，因为乐高在某种层面上是有一点儿抽象的。你不能用很多小鸟来制造大鸟。即便你做到了，那也不是一只真实的鸟。但是你可以用很多小鸟模型来制作一个大鸟模型，因为鸟的模型足够抽象，而真实的鸟并不是抽象概念。有艺术家将小的人物肖像作为马赛克砖，用这些肖像马赛克拼成一个大的人物肖像。你可以用肖像来制作肖像，因为肖像足够抽象，而真实的人类则不是抽象概念。

当我还很小时，我曾学习过在频谱计算机上编程，就像一整代英国程序

员、数学家以及其他这类人士的家庭一样。我已经谈论过我最喜欢的能打印输出无穷“你好”的程序。我不厌其烦地把频谱计算机的概念描述给那些没有见过频谱计算机的人，因为它如此简单，又如此非凡。以防你没听说过，我现在先介绍一下频谱计算机。它可能是家用计算机最早的版本之一。它基本上只由一个键盘组成，那是一个很小的带有橡胶按键的装置。它没有屏幕，你需要把它接入你的电视接收器。你需要使用传统的盒式磁带和磁带记录器来保存信息。你只需在机器上按下“记录”按钮（也就是录音键），并且敲击频谱计算机上的“保存”按钮，它就会将你的程序记录成一段有趣的小调。通过哼唱这些小调，我对它们有着非常鲜明的记忆。（每一个伴随着频谱计算机成长起来的人都会被这段记忆打动，可能还会发出一阵咯咯的笑声。）要再一次加载某个程序，你需要按下磁带记录器的播放键，敲击频谱计算机上的“加载”按钮，它就会重播那些有趣的音调，并神奇地加载出程序。你也可以购买盒式磁带，商家会在上面装载许多令人兴奋的游戏。

当时的打印机也是一个非常可爱的小东西，我已经忘记如何把频谱计算机和它连接起来了。真正令人难忘的部分是它使用的银纸卷，它就像厕纸一样是卷起来的，但是更小。打印的那一面含银，我认为它是使用加热的方式来成像的。那时候，我们只有一种字体和一个字号，而且字符之间的间隔是完全相等的，就像Courier字体一样。

我喜欢使用这个打印机来制作一些标志，比如“尤金妮娅的房间”（EUGENIA’S ROOM）。但是因为无法更改字号，所以为了使标志足够大，我需要用字母本身来组成每一个字母，这件事特别有趣。就像这样：

EEEEEEEEEEE

EEEEEEEEEEE

EEEE

EEEEEEEE

EEEEEEEE

EEEE

EEEEEEEEEEE

EEEEEEEEEEE

以此类推。字母可以迭代，因为它们也是抽象的事物。数学中的所有事物都是抽象的，因此你会经常用已有的数学概念来构建更多的新的数学概念。例如，你可以用A来画E，虽然这有点反直觉。

AAAAAAAAAA

AAAAAAAAAA

AAA

AAAAAA

AAAAAA

AAA

AAAAAAAAAA

AAAAAAAAAA

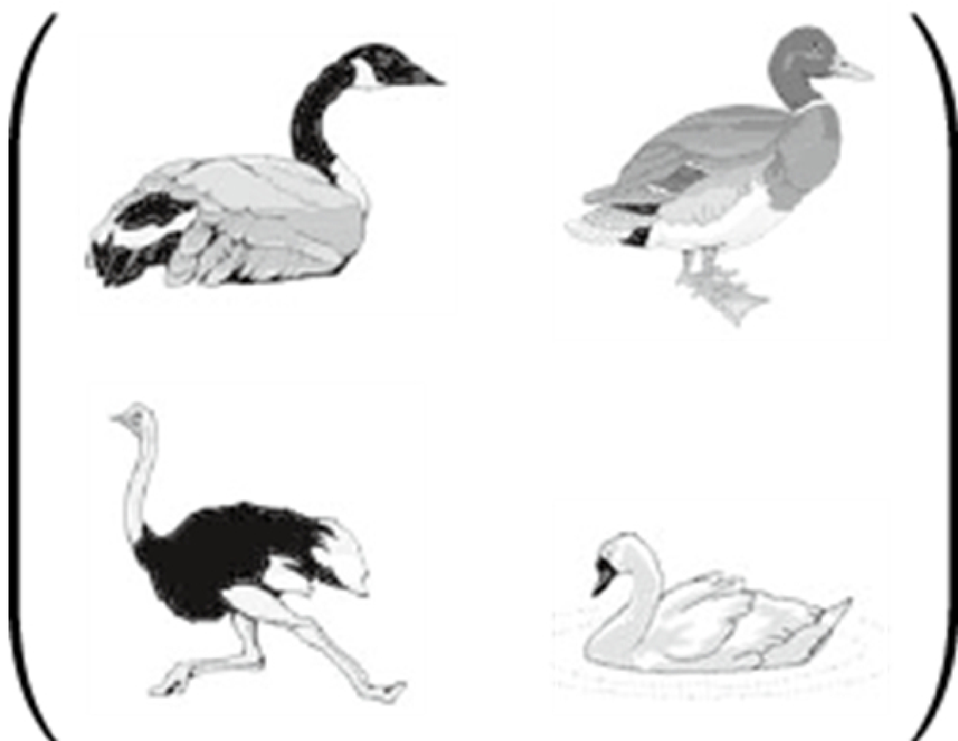
你可以用数学对象来做类似的构建工作，因为它们也是抽象的。例如，你可能还记得矩阵是什么。它们是这样的：

$$\begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix}$$

这是一个由数字组成的矩阵，但是我们也可以用其他事物来组成矩阵。当我在学习初级高数时，图形计算器刚刚问世。神奇的是它们居然能处理矩阵。初级高数考试的出题人很快意识到了这一点，于是用字母取代了数字，挫败了我们使用计算器考试的意图。由字母组成的矩阵类似这样：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

但我们也完全可以用禽类组成一个矩阵：



或者用矩阵来组成矩阵：

$$\begin{pmatrix} \begin{pmatrix} 1 & 0 \\ 3 & 2 \end{pmatrix} & \begin{pmatrix} 2 & 5 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 3 & 2 \\ 0 & 1 \end{pmatrix} & \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} \end{pmatrix}$$

矩阵是足够抽象的，因此我们可以用矩阵来制作矩阵。数学概念是足够抽象的，所以我们也可以用数学概念建构更多的数学概念。克里斯托弗·丹尼尔森在他那本奇妙的著作《哪一个不属于？》中举了一个由多个“哪一个

不属于”的例子组成的“哪一个不属于”的例子：哪一个“哪一个不属于”不属于这里？

在我的前一本书中，我谈论了我的研究领域——范畴论。范畴论是“数学的数学”。那范畴论的数学又是什么的呢？还是范畴论，但是它包含更多的维度。为什么会这样？因为范畴论是关于事物间关系的理论。那么，如果这个“事物”本身就代表一种关系呢？

飞机、火车和汽车

在这本书的开头，我比较了乘飞机旅行和乘船旅行。如果我乘飞机越过大西洋，我的目的不会是看风景，因为旅途的大多数时间并没有什么可看的。也许有一天，我会乘船横渡大西洋，但通常来说，我都希望能更快地到达目的地。

当你去度假时，你是会考虑不同的交通方式，还是总是乘坐飞机？也许你决定去阳光明媚的地方，在这种情况下，如果你住在英国，你就很可能需要乘飞机，不然你就要坐很长一段时间的火车。

小时候，我家常常在假期去法国，参观葡萄园、买酒。我们住在布赖顿附近，开车到渡口很方便。一般来说我们有两种选择：开车到纽黑文，然后坐船到迪耶普；或者开车到多佛，然后坐船到加来。开车去纽黑文的距离非常短，但从这里横渡到法国迪耶普的距离就长得多。我不知道哪一个更贵，当时我还太小，问不出这样的问题。

当我还非常小的时候，气垫船也是一种可选择的交通方式。这个神奇的装置看起来像是能够飞过水面似的。可以想见，我第一次乘坐气垫船时是多么失望，我发现我根本不像是在飞行，我的感觉和坐在一艘船上一模一样，而且根据我的记忆，它更易受到海浪的影响。结果就是，我们全都严重晕船了。

当你有了小孩，开车出行是一个更可行的选择。除了便宜外，你还可以把所有的育婴用品都放在车里，并且仍能剩下充足的空间放购物袋，特别是如果你的车像我家的老萨博一样有个很大的后备厢的话。如果你决定开车到某处，你可能还需要比较不同的路线。如果你在谷歌上搜索行驶路线，它会给你提供一些选项。你可以比较驾驶的时间、距离和道路类型。如果你像我一样，你很可能会选择最简单的路线，走主干道以避免迷路。不过，也有很多人会选择最短或者用时最少的路线。

所有这些都是在比较从一个地点到另一个地点的路线。从A到B的路线是这两个地点间关系的一种表现形式，而且对于同样的两个地点，我们可以找

到很多不同的路线。如果现在我们观察不同路线间的关系，那就是在观察不同关系之间的关系。

范畴论研究的正是事物之间的关系，现在我们在范畴论领域得出的维度概念如下所示：

- 如果你忽视事物之间的关系，假设一切都在真空中，那就是零维。

- 如果你接受存在事物之间的关系或者两个地点间的路线，那就是一维。

- 如果你考虑这些关系之间的关系，那就是二维。

- 如果你考虑这些关系的关系之间的关系，那就是三维。

……以此类推。

这里我想说的是，如果我们已经决定考虑事物之间的关系了，那么我们为什么不也考虑一下这些关系之间的关系呢？研究事物之间的关系是要将事物放在环境中。如果确实如此，难道我们不该在环境中研究关系吗？我们可能会比较去度假的不同方式，发现乘飞机更快而开车不那么贵。接下来我们可以比较这些关系：速度是否比花费更重要？或者在现实中你可能会八卦其他人之间的关系（当然最好不要）。你可能发现一对伴侣在一起非常快乐，但也会发生很多争吵，而另一对伴侣从来没有争吵过，但是他们看起来似乎也不太快乐。那么你就要对比是快乐比较重要，还是没有争吵比较重要。

范畴论研究事物之间的关系，并以各种方式在这些关系的基础上延伸：根据事物的性质来表征它们，寻找事物能够发挥最大价值的空间，将事物放在环境中，表达事物“或多或少相同”这个微妙的概念。更高维度的研究也全都是建立在关系本身上的。这是另一个级别的抽象，它将我们带入更高维度的范畴论。

在范畴论中，我们以箭头表明事物间的关系，如图13–3所示：

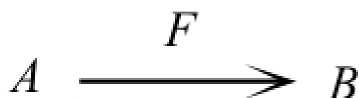


图13–3

A是我们居住的地方，B是我们要去度假的地方，F是我们到达那里的方式。我们的旅行就是家和度假目的地之间的“关系”。现在，我们有两种可行的方式从家前往目的地——F和G，以及一个比较它们的方式。在更高维度的范畴论中，我们可以把它画成这样的示意图（见图13–4）：

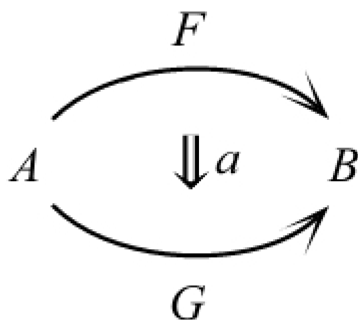


图13–4

这里标记为a的双箭头表明的是路线F和路线G之间的关系。这个关系可以包含关于花费、时间、乐趣或沿途风景的比较。

我们也可以像这样画出如图13–5所示的大图，这是来自我的研究项目中的一个真实图例。

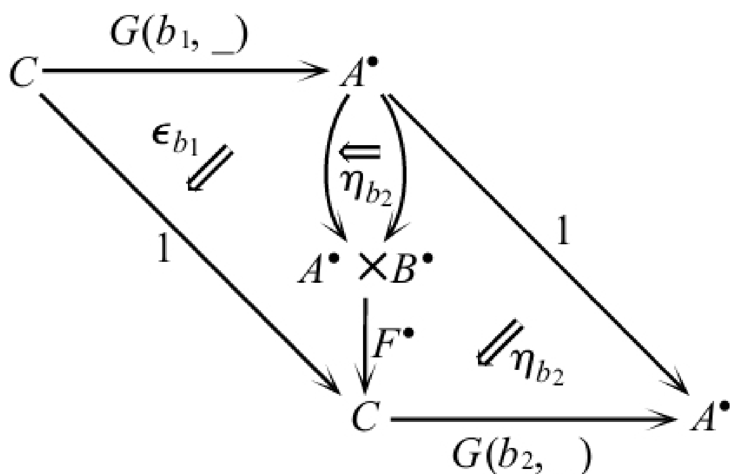


图13–5

高维范畴论就是在研究这些更高维度的结构，它们会不停地增长，直到你喊停，并表示你已经获得了足够多的维度。但是如果你不喊停，它们将会持续增长到永远。这样，我们就获得了“无穷维度类别”。接下来我们会发现，从某种意义上说无穷维度的类别会比有限维度类别还简单些，因为它们的增长是更加“自然的”。在数学里，特别是在范畴论里，没有人为限制的有机增长的事物往往不会太复杂，而我们强加的人工制约虽然能够使事物的结构变得更加实用，但是也会让它变得更加复杂。在某种程度上，这就像一个关于人类社会的事实：成人世界比起儿童世界在想象力和理想主义方面会有更多的制约。这些制约使得社会更加实用化，但是也变得更加难以捉摸。当你不必将它转化成一些可行的逻辑的时候，理想主义更容易理解。

在范畴论里，理想主义和逻辑学之间总是存在着一种张力。有许多结构自然地想要无穷的维度，但是这太不实用了。所以我们在一个有限维度的环境中研究并限制它们，并且与为了使这些结构在逻辑上可行所产生的后果做斗争。在上一章中，当我们在考虑评估一个情况的标准时，我们看到了不同的降低维度的方法。类似地，范畴论中也有不同的方法用以降低维度数。每一种方法都会引起不同的问题。你可能想知道，为什么我们不能像上一章那样直接忽略一些事物？在了解了数学如何演绎之后，我们会回到这个问题上。我们考虑的每一个维度上的关系都会创造出一个新的需要我们研究的关系的维度。类似地，我们回答的每一个问题都会引发更多的问题。

攀登每一座山

电影《我们要活着回去》（*Alive*）讲述了关于1972年发生在安第斯山脉的乌拉圭空军571号班机空难的可怕故事。飞机在恶劣天气下撞上了山腰。在电影中，我们看到幸存者们接收到了一个无线电信号，听到一条新闻报道说对幸存者的搜救行动已被取消。于是他们决定派一小队人走出山脉寻求帮助。整个乌拉圭的橄榄球队都在飞机上，这部分幸存者看起来非常适合这项任务。

他们中的三人带着他们认为足够保障其安全的物资出发了。在下到山谷前，他们必须首先越过视野内不远处的一座山。但是当他们到达那座山的顶峰时，他们看到的是横亘在他们面前的更多更高的山脉。之前，更高的山脉被山脊遮挡住了，但是现在他们爬得足够高了，他们看见了真正需要越过的山脉。

这时他们意识到，他们没有足够三个人存活的物资，所以他们中的一人交出了他的用品回到了撞机地点等待救援。情况困难得不可想象，但是他们确实走出了山谷，留下的幸存者们也获救了。

数学没有这个故事这么可怕，但它确实像一道无休无止的山脉。每当你征服了一个高峰，产生了片刻的喜悦之后，你就会发现征服这座山峰刚好令你能够看见前面更高的山峰。这有一点儿像我在新墨西哥爬山时爬到峰顶，接下来却只看见面前绵延的山脉，或者我沿着弯曲的湖岸游泳时，每次靠岸，我只能看到更长的湖岸线。

这是不可避免的。事实上，这正展现了数学的力量。因为我们所学的概念往往可以用来构建更大的概念，而新的概念又能够用来构建更大的概念，以此类推。这些都源于它的抽象本性。如果我们正在登山，我们并没有用已经攀登的山脉建造更多山脉，我们只是看见了更多原本就存在的山脉。但在数学中，我们要持续构建越来越大的概念。

我们并不是故意那么做的，而是碰巧我们开发出来的研究数学对象的方法就是数学的新的组成部分。而为了研究这一新的组成部分，我们又创造出了需要研究的数学的更新的组成部分。这在你研究禽类时是不会发生的，因为你开发出来的研究禽类的方法本身并不是禽类。

这就是范畴论产生的基础，它就是那个研究数学的新的数学组成部分。在某种程度上，范畴论是终极的抽象。你利用科学来抽象地研究世界，利用数学来抽象地研究科学，利用范畴论来抽象地研究数学。每一个阶段都是进一步的抽象。但是要抽象地研究范畴论，你还是要利用范畴论。

总是崩溃的计算机

你可能想知道，是不是直接在我们的理论中添加维度就能解决问题了？这是什么意思呢？为了在二维空间中明确一个点的位置，我们需要给出两个坐标：一个x坐标和一个y坐标。为了在三维空间中明确一个点的位置，我们需要再添加另外一个坐标。我们可以根据我们的需要添加更多的坐标来构建我们想要的维度，即使我们并不知道在“空间”中这些维度看起来是什么样的。

坐标、维度都不必然涉及物理空间。只要我们有四个独立变量，我们就有四个维度。它们并不非得是空间或时间维度。就像我们上一章所看到的，维度观念可以产生于我们用来评估事物的标准，也可以产生于机械臂的铰链。

我们可以用这种方式来研究计算机崩溃。与物理空间不同，我们可以用计算机内置的配置组合来构建抽象空间。如果一个路径最终通向死胡同，那么就会导致崩溃。拓扑学是研究空间的一般形状的数学分支，我们也可以用其来研究抽象空间。在拓扑学中，环形甜甜圈与带有把手的咖啡杯具有相同的一般形状。因为它们都只有一个洞——在拓扑学中，我们基本上只考虑这个特性。

既然如此，为什么这一切还是那么困难？唉，无论我们尽多大的努力，计算机还是会崩溃。原因是，很不巧，它比我们所描述的要更微妙一些。因为我们在增加维度的时候，增加的不仅仅是关系间的关系，我们还增加了这些关系所遵循的理论或规则。我们添加的维度越多，这些原理就越难理解。

增加维度带来的问题之一就是“结合律”的问题。在最常见的情况下，结合律意味着：

$$(3+5)+5=3+(5+5)$$

换言之，我们在计算中将括号放在哪里都无所谓。这非常方便，但是在处理比数字更加复杂的事物的时候，这个公式并不总是正确的。我先前曾讲过关于鸡蛋、糖和牛奶的例子。在这个例子里，结合律就是错误的：

（蛋黄＋糖）＋牛奶

能够制作出蛋奶冻，但是，

蛋黄＋（糖＋牛奶）

并不能制作出蛋奶冻。

这里也有一个数学上的例子。拓扑学研究空间形状，代数拓扑学则通过观察空间中的行程来研究它。上面两句话就涉及范畴这个概念。范畴是一个数学结构，它具有：

* 对象

* 对象间的一些关系，用对象间的箭头表示

* 如果箭头指向相同方向，那么这两种关系就可以合并。所以如果你有如图13–6所示的箭头：

$$A \xrightarrow{f} B \xrightarrow{g} C$$

图13–6

你可以将它们合并成如图13–7所示的图形：

$$\mathrm{A} \xrightarrow{g \circ f} \mathrm{C}$$

图13–7

在上面的例子中，我们的对象是空间中所有可能的点，我们的箭头是从一个点到另一个点的行程。我们能够结合它们，是因为如果你有一个从A到B的可行的行程和一个从B到C的可行的行程，那么你就可以将两个行程连起来获得一个从A到C的更长的行程。

通常当我们考虑从A到B的行程时，我们也会考虑这会花费多长时间。这个时间可能是以分钟、小时或者其他标准计的。在抽象数学空间中，我们并

没有真正的距离单位，此外我们也并没有真正的时间单位。距离和时间都只是数字。如果你像我一样，是那种常常因为在学校测试中忘记写单位而被扣分的人，你将会为这次我们不需要单位而感到欣慰。

为了便于比较行程，我们假定所有的行程都会花费时间1。这就像当我们使用百分比时，我们知道所有的分母都是100。因为这样一来我们就能更容易地比较事物。当我们将总时间“标准化”为1时，我们便只需考虑选择哪条路径，以及相对于整体行程，每一段行程需要花费多长时间。

我们知道，如果我们有一个从A到B的行程和另一个从B到C的行程，我们可以将它们结合在一起获得从A到C的行程。问题是如果我们将它们结合在一起，时间的总长就会变成2，而不是1。因此，为了“标准化”这段行程，我们现在假设每一段行程都以两倍的速度进行。因此，现在我们花费了相当于原来一半的时间从A到B，再从B到C。

这很好。但是当我们试图合并三段行程时，问题就出现了（见图13–8）。

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$$

图13–8

你也许会想，将每段行程以三倍速度进行就可以了。确实可以，但这不是结合律工作的模式。我们不能直接发明一个新方法将三个事物结合在一起。我们必须使用原始的方法，通过在不同地方使用括号先将两个事物结合在一起，看看会发生什么。换句话说，如果我们首先结合下面两段行程（见图13–9）：

$$A \xrightarrow{f} B \xrightarrow{g} C$$

图13–9

再在末端连上h，这种做法应该与首先结合下面两段行程（见图13–10）：

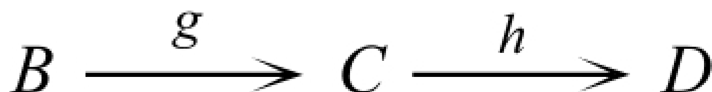


图13–10

再在前端连上f的结果是一样的。问题是，这并不正确。因为在第一串路径中，我们将f和g结合在一起而每一部分以两倍速度进行，所以我们完成它们会花费如下时间量：

行程	时间花费
f	$\frac{1}{2}$
g	$\frac{1}{2}$

然后我们将h结合在末端，我们同样需要将每件事以两倍速度完成。所以我们需要花费总时间的一半来做f和g，另外一半来做h。这意味着我们需要花费如下时间量：

行程	时间花费
f	$\frac{1}{4}$
g	$\frac{1}{4}$
h	$\frac{1}{2}$

但是，如果我们在开始的时候就将行程g和h先结合在一起，每一部分都以两倍速度完成。那么它们会花费的时间量为：

行程	时间花费
g	$\frac{1}{2}$
h	$\frac{1}{2}$

接下来，我们将f结合在前端，同样，我们将花费总时间的一半来做f，另一半来做g和h，也就是花费如下时间量：

行程	时间花费
f	$\frac{1}{2}$
g	$\frac{1}{4}$
h	$\frac{1}{4}$

比较第二个和第四个行程时间花费表就会发现，我们确实没有得到相同的答案。在第一串路径中，我们在f和g上分别花费了$\frac{1}{4}$的时间，而在h上花费$\frac{1}{2}$的时间。但在第二串路径中，我们在g上花费了$\frac{1}{2}$的时间，而在g和h上分别花费了$\frac{1}{4}$的时间。因此这种结合方法是不符合结合律的。换言之，我们放置括号的位置确实会产生影响。

数学里处理这个问题的方法有很多，具体选择哪种取决于你希望在现实逻辑和抽象理想之间取得怎样的平衡。近似处理可以让我们获得更多的可行性，但是会造成一些精确性的丢失。一种处理方式是将一个更强的“相同”概念施加在这个行程上，即将这两个你花费不同时间量的版本计为相同。这通向一个被称为同伦论的研究领域，这里涉及的“相同”这一概念被称为同伦。

另一种处理方法是用树状图来持续跟踪这些细微的差异，考虑括号的所有可能的位置。图13–11中的两个树状图分别描述了括号位置不同的两种情况。

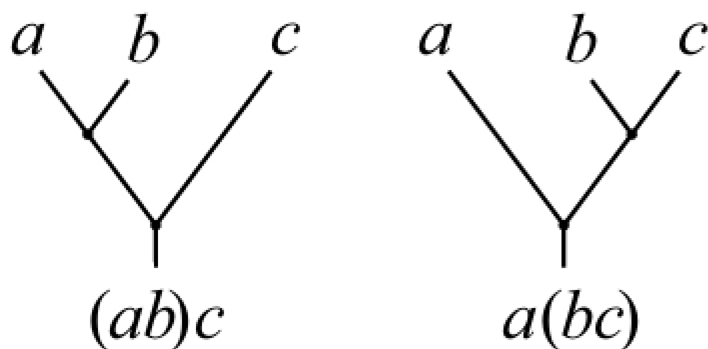


图13–11

这通向了一个被称为Operad理论注的研究领域，因为树状图形成了一个被称为Operad的代数结构。

另一个处理方法是观察这些行程间的关系，这通向了高维范畴论。随着维度增加，一些微妙的事发生了。我们可以继续观察这些行程（在数学上通常被称为“路径”）之间的关系，以及这些关系之间的关系，以此类推。我们可以在任何一个时间点决定停止考虑更高的维度。然后我们会在那个最高维度处施加一个“相同”的概念，这样我们就不必再考虑更多的维度了。这有点儿像决定哪些特定标准不是那么重要，然后你就可以忽略它们一样。简单来说，即使一些对象原来在这些标准上存在不同，现在我们也认为这些对象是相同的了。例如，你可以决定在买鸡蛋的时候不关心鸡蛋是棕色的还是白色的。

在范畴论中这么做的问题是，我们需要考虑这个做法对于原理产生的影响。事实证明，某种程度上，不停思考更多的维度反倒更容易些。因为这么做的话，当事物并不真正相同时，我们不用强制认为它们“相同”并处理由此引发的后果。

这听起来可能与直觉相反，但是用下面的方式思考一下：如果我们决定，应该有一个最大的数字。比如，我们说，“就这样吧，大数字太复杂了，从现在开始没有比1000更大的数字了”。这将引起可怕的后果，因为它是违背自然的。我们确实“需要”无穷的越来越大的数字，即使我们从来不使用这些数字。

对于范畴论中的维度，我喜欢这样考虑它：每增加一个维度，就相当于添加了一个新的体现事物微妙差别的层面，这样，我们就能把解决由强迫事物相同导致的对某条原理的违背所产生的问题再往后推迟一点儿。如果我们有无穷的维度数，我们就能永远推迟这件令人不悦的工作。如果我们长生不老，我们就能永远推迟下去了。

---

1. Operad理论是一个探讨典型代数问题的抽象代数领域，例如探讨可交换性或反交换性，以及各种结合性。——编者注

14　无穷小

我依然会为乘飞机到达一个城市并从空中欣赏这个城市的建筑而感到兴奋。近看那么高大的建筑物，从飞机上看却显得那么小，这仍旧让我感到神奇。曼哈顿对我来说格外超现实，因为那么多的摩天大楼竟然可以全部塞进一个小小的岛屿。飞往香港启德机场的老航线则异常可怕，因为飞机飞得离建筑物非常近。这一分钟摩天大楼还小得像芽尖，下一分钟你就能看到某人的公寓，甚至看见住客在里面洗衣服。

迄今为止，我们观察的大部分对象都涉及事物的增长和无穷的事物的出现，因为前者总是会引发后者。但是现在，我们要快速跨过这个阶段，转而考虑非常小的事物。通过研究无穷多个无穷小的小事物，我们可以从一个完全不同的角度来获得无穷多的事物。这看起来是不是有点儿像作弊？有时，数学的进步通过稍稍转变看待事物的视角就能实现。这并不意味着构建新事物或者前往一个完全不同的领域，只意味着改变你的观点，而由此你就很可能开启了一个巨大的、新的可能性。这里，无穷多的无穷小的小事物将我们引向微积分，由此，我们得以开始理解弯曲的事物、运动中的事物、流动的事物以及持续变化的事物。

我们的世界中很少有事物不符合上面这个描述。大部分计算机是数字的，也就是说这些计算机的功能可以被分解为精确定义的数字以精确定义的步骤发生变化，而不是连续变化。但是计算机运行所需要的电还是属于会“连续变化”的领域。现在我能看到的少数不涉及微积分的物体就是我正在使用的桌子。确实，桌子在微积分发现之前就存在很长时间了。但是我用的这个桌子是宜家工厂生产的，整个生产过程肯定会涉及微积分。我的观点是，的确，对于无穷的研究从字面意义和真实意义上讲都是抽象的，但是它会将我们带入微积分的领域，而微积分与现代生活的方方面面都密不可分。

所有这一切的出发点是考虑“无穷紧密连接”的事物。当我们在电脑上用画图工具画一个圆的时候或者输入字母“o”的时候，它看起来是很平滑的，且首尾相连。但是如果我们把它放大足够多的倍数，它最终就会变得像素化。图14–1是在我电脑屏幕上被放大数倍的字母“o”。

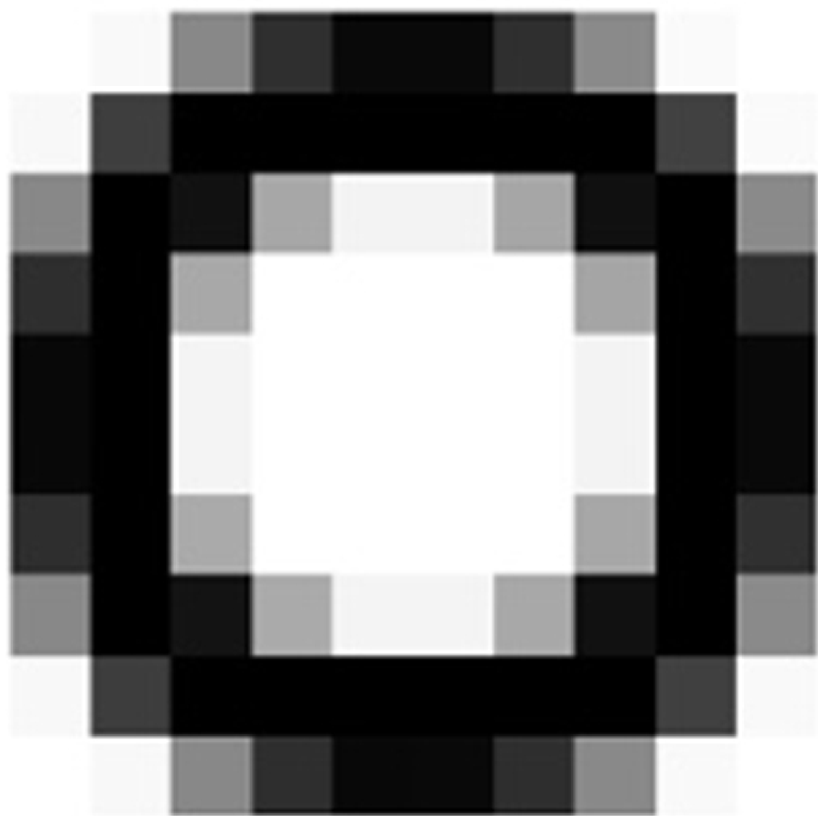


图14–1

我们能看到，这个图像只是有限数量的方块通过拼接伪装成的一条首尾相连的曲线。为了更好地伪装曲线，我电脑中的这幅图还包含一些灰色的阴影。为了显示图像，计算机只能这么做，因为它只能理解独立的点。它只有有限数量的点可以用，而且这些点都具有固定的可测尺寸。

但是我们的大脑呢？从微积分的意义上说，我们的大脑能做得更好，因为我们可以处理无穷多的事物，而且我们可以在这些事物无穷小的情况下对其进行处理。这就是我们现在将要研究的。

我曾经在剑桥的派克街小学帮助孩子们补习数学。有一次，我试图帮助两个6岁的孩子理解对称性。我让他们在一些三角形上画对称线，接下来在正方形、正五边形、正六边形上画对称线。最终我给了他们一个圆。他们

中的一人画了这条线（见图14–2）：

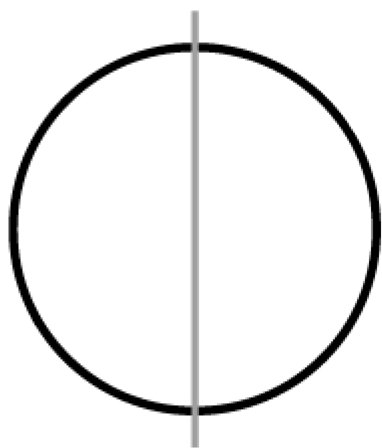


图14–2

接下来，另一人画了这条线（见图14–3）：

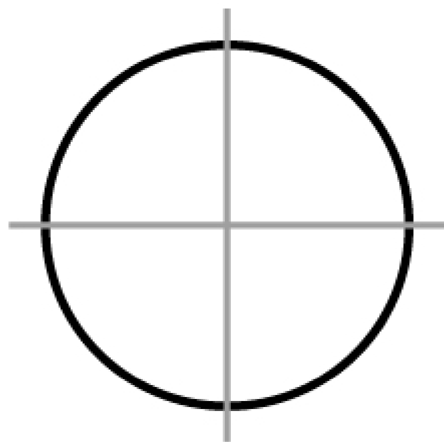


图14–3

然后一人又画了另外两条线（见图14–4）：

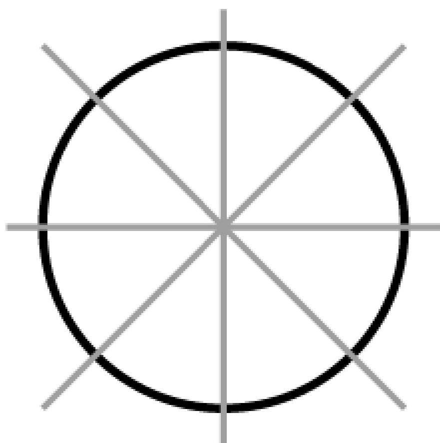


图14–4

后来，事情变得非常激动人心。第一个孩子叫道，“圆有几百条对称线！”接下来第二个孩子说，“圆有几百万条对称线！”然后，第一个孩子又叫道，“你要用整个人生来画它们，你永远也画不完！”停顿了一会儿之后，另一个孩子拿起一支铅笔，涂满了整个圆，然后说，“看吧，我完成啦！”

我感到很难办。我不得不承认，他们都是对的。你可以用一生来画圆的对称线，却永远无法画完，因为它们有无穷多，而且事实上是不可数的无穷多。一种理解这件事的方式是通过说出对称线与水平线间的夹角来明确它的位置（见图14–5）。

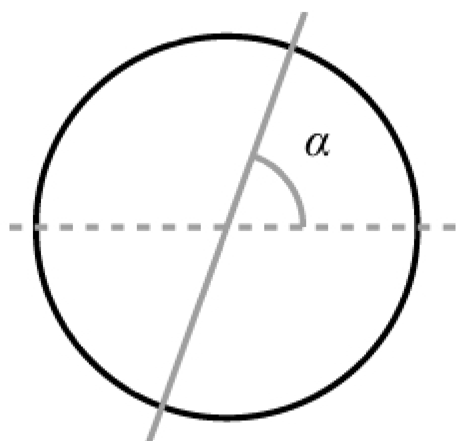


图14–5

对于α，我们可以选取0～180度间的任意角度，或者0～π之间的任意弧度。任何α大于180度或者π的线都会和已经画过的线重叠（见图14–6）。

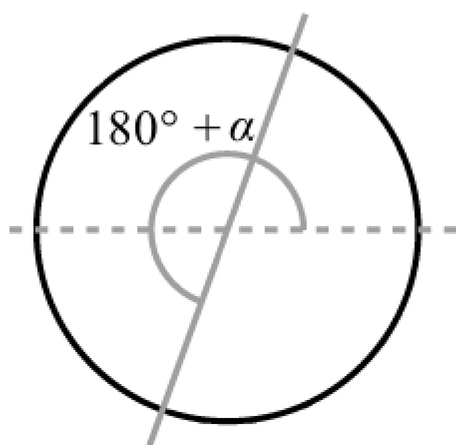


图14–6

对于α的数值，我们可以选取1到180之间的任意实数，而不仅仅是整数或者有理数。我们已经知道，光是在0到1这个区间内就有无数个实数，所以

我们可以确定地说，在0和180之间也有无数个实数。

所以对于一个圆来说，它确实有无数条对称线。但是如果你涂满整个圆，你也就确实把它们全画出来了。你也许会认为这是作弊，因为在圆的中心，真实的对称线会重叠无穷多次，所以你其实应该在中心涂无穷多层。但是如果我们不考虑中心，只是尝试沿着圆的边缘标记所有的对称线，那么我们沿着圆的边缘描上一圈就可以简单标记出所有的对称线了。在这个过程中我们画出了无穷多的点吗？换句话说，下面这条线上有无穷多的点吗（见图14–7）？



图14–7

如果有无穷多的点，它们有多宽？如果只有有限多，那点有多少个呢？

除以无穷

如果我们把上面那条线分割为越来越多的线段，线段肯定会越来越短。那么，我们能将它分割成无穷多的小点吗？这个问题的关键在于，我们能否通过将某物无穷分割来制造无穷小的事物。

想象一下，你有一张彩票，其号码组合可以是所有的自然数，即每一个自然数都会产生一种可能的结果。这意味着，彩票机中有无穷数量的号码球，每个小球上都写有一个有限的数字。如果真的是这样，那中奖的机会将非常渺茫。常规的英国彩票是在59个小球中选取6个。数字组合的可能性大约有4500万种，所有这些组合都具有均等的可能性，所以你获胜的概率大约是四千五百万分之一，约等于0.00000002。这是一个非常小的数字，但它并不是零。虽然在我看来它与零是那么接近，而在现实中这个数也几乎被当作零来处理了。当你用它与可能的数字组合的总数（4500万）相乘，你会得到1。这是应该的，这等于是在说，如果你买下所有可能的数字组合的彩票，你赢的可能性就是100%。

有无穷个小球的彩票机将产生无穷多的组合，所以你中奖的概率应该是“无穷分之一”。作为分数，这个概率意味着什么？首先，它不可能等于比零大的数，因为如果它比零大，那么当我们将它与可能的组合总数（无穷）相乘，我们将得到大于一的数。那么，这是否意味着这个概率是零

呢？但是某些人确实每次都能中奖。此时，你可能会争辩说，在实际中不可能有这种彩票。但是这就和在希尔伯特旅馆实验中说这样的旅馆是不存在的差不多。换言之，这种说法并不能真正解决矛盾。

这将我们带回到我们第一次尝试定义无穷的时候，那时我们尝试假设：

$$\frac{1}{0}=\infty$$

我们已经知道，在等式两边同时乘以0，将会产生一个矛盾的结果。与此相关的另一个假设是，我们可以尝试用1除以无穷来得到0。也就是，

$$\frac{1}{\infty}=0$$

现在我们已经对无穷有了更多的了解，我们马上就会发现这个等式也有错误之处。现在的问题是，我们之前定义无穷的方式是使用事物的无穷集合，其中并没有涉及任何关于除以无穷的概念。对于这个问题，数学上较好的回应方式是，“好吧，我们接下来就试一试”。我们以前没有做过并不意味着它是不可行的。

我们可以采取和研究减法类似的方法来研究这个问题，再次将一切事物都考虑成对象的集合。这就像数积木一样，你不能切碎积木。如果我们正在考虑自然数的集合，我们就不能将它们切碎成部分自然数。

当我们定义与无穷相关的减法时，我们不得不回到原点，将6–3考虑为“从3数到6我需要数几个数”。也就是说，我们考虑用以下等式来解决问题：

$$3+x=6$$

现在我们考虑做$6\div3$。我们可以用两种方式来考虑。

- 叠加几次3能得到6？换言之，我需要将3与它自身相加几次能得到6？这与解下面的等式是一样的：

$$3\times x=6$$

- 哪个数字叠加3次能正好得到6？换言之，哪一个数字与它自身相加3次能得到6？这与解下面的等式是一样的：

$$x \times 3 = 6$$

两种情况下的答案都是2。因为对于有限的数字来说，变换乘数和被乘数的位置并不会带来什么影响。但是当涉及无穷时，我们会发现这两种情况并不相同。

例如，无穷次叠加3与叠加无穷3次是不同的。也就是，

$$3 \times \omega \neq \omega \times 3$$

当我们问“我需要将3与它自身相加几次能得到ω”时，答案是ω。想象一下，你又成了给队列中的客人发放票据的人。这一次，人们三人一组地到达。在你使用完一个无穷票据本之前，你给多少个三人组发放了票据？答案是ω，因为你要每次分发三张票据直到永远。

但是，如果我们反过来问“哪一个数字与它自身相加3次能得到ω”，这时就不存在一个可能的答案了。如果你将3个有限数字相加，答案也将是有限的。如果你将3个无穷数字相加，那么它们每一个都至少与ω一样大。因为ω是最小的无穷，所以将3个ω相加的结果应该会更大，就像“永远还要再多一天”比永远要久一样。我们可以再一次通过发放票据的例子来考虑这个问题。如果现在有无穷位客人同时到达，你至少会用完一整本票据。那么当下一批客人到达时，你注定需要另一本不同颜色的票据。

这两个问题都是在尝试做“无穷除以3”，但它们得到了不同的答案。这表明在考虑无穷时，除法与减法一样，不是一个很好理解的概念。而且这只是无穷除以一个小的有限数字的情况。如果我们试图除以无穷，结果会更糟。假设我们要做“$\frac{1}{\infty}$”，或者说“1除以ω”，这时也会有两个可能的问题：一个是，“我需要将ω与它自身相加几次能得到1？”这肯定是不可能有结果的，因为从一开始，ω就太大了。另一个是，“哪个数字与它自身相加ω次能得到1？”同样地，这个问题也没有可能的答案。

尽管如此，$\frac{1}{\infty}$看起来确实应该是零。对于上面两个问题，这会是个明智的答案吗？如果我们将ω与自身相加0次，我们得不到任何东西，因此这是没有意义的。这就像有零批无穷的客人到达一样，事实上，你根本不需要用到任何票据。对于第二个问题，我们将0与其自身相加ω次能得到1吗？这就像零个人到达队伍无穷数次一样，你还是不需要用到任何票据。

此时，我们可能会放弃地说“好吧，$\frac{1}{\infty}$不是零”。但我们也许可以做些更符合数学特点的事，比如，“如果$\frac{1}{\infty}$等于零真的有其合理性，那么我们能不能不以先前思考无穷数集的方式为基础而赋予这个表达式一些其他的数学意义呢？”数学要做的事之一就是找到那些我们直觉上看起来是对的事物，然后给予其一个精确的符合逻辑的解释。我们不应该这么轻易放弃。

无穷的对立面

此刻，你可能想知道我们能否发明一个不为零的、无穷小的小事物。因为我曾说过，我们可以通过想象来使抽象的事物存在。数学家的确尝试过这么做。因为这貌似是有道理的，就像在我们认真思考无穷这个概念之前也觉得无穷貌似是有道理的一样。这个概念就像是无穷的对立面：无穷比所有数字都大；无穷小则比所有数字都小。如果你将无穷与其自身相加，你会得到无穷；如果你将无穷小与其自身相加，你会再次得到无穷小。此外，当你将无穷与无穷小相乘，你会得到1，如此你就能解决彩票中奖概率的问题了。

通过这种方法构建无穷小会产生的问题和仅仅通过“想象”构建出无穷会碰到的问题是一样的。这一概念的建构能够通过认真的思考，或者更准确地说，通过高超的技术来实现。但是通常来说，绕过问题是更容易而且更巧妙的。如果你散步的时候遇到一个大泥坑，你可以就这么踩进去然后寄希望于你的靴子足够防水，你也可以绕过它继续往前走。（当然，许多人，尤其是孩子，喜欢故意直直地走进泥坑里去。在数学领域中也常常有人这么做。）

以下是我们巧妙地避开除以无穷这个问题的方式。想象一下，你把一整个巧克力蛋糕分享给一群人。如果只有2个人分蛋糕，那每个人都能分得很多蛋糕。如果有3个人分蛋糕，每个人仍然可以得到很多，但是比之前要少。如果是4个人，每人得到的会更少。分蛋糕的人越多，每人分到的蛋糕就越少。如果分蛋糕的人数非常多，尝试分享一个微不足道的蛋糕就变得非常愚蠢了。你是否曾经把一个蛋糕分给100个人？（结婚蛋糕不在这个行列，因为它通常有好几层，是真正的多层蛋糕。）或者分给1000个人？100万人？当分蛋糕的人太多的时候，每个人都只能获得非常小的一部分了。此时每个人分到的蛋糕的量是如此微不足道，基本上就等于什么都没有。

当有100万个人和一个蛋糕的时候，严格意义上讲每个人还是能分到一些

蛋糕的，其中甚至可能包含几亿个分子。但是每个人分到的部分看起来近似于零。而且随着人数的增加，它会变得越来越接近零。这就是我们赋予“除以无穷为零”一个数学含义的方式。我们通常不会除以无穷，因为这么做并不具备明确的意义。就像在第11章中一样，我们可以回归到接近无穷的事物这个想法。我们除以某个接近无穷的事物，然后我们发现答案接近零。一些自作聪明的人可能会带上一个显微镜，然后说他们仍能看见手里的蛋糕。但是我们可以继续把蛋糕分下去，这样它就又变得不可见了。这并不意味着1除以无穷“等于”零，但是它提供给我们一个方法，让数学可以赋予这个直觉观点一些含义，而这正是整个现代微积分的开端。

芝诺悖论

微积分的概念可以追溯到很久远以前。早在2500年前，希腊哲学家芝诺就研究过关于由无穷多的无穷小部分组成的连续体的一些难题。就像几千年后的希尔伯特一样，芝诺认为，有些悖论提醒我们，在思考无穷事物的时候需要格外小心。

芝诺悖论之一所描述的情形和小孩子们吃美味的巧克力蛋糕时会思考的问题差不多：如果你吃掉蛋糕剩下的一半，然后再吃剩下来的一半，持续吃每次剩下的一半，这是否意味着你的巧克力蛋糕会永远吃不完？

芝诺则是这样表达的：如果你想从A移动到B，首先你需要移动一半的距离，接下来你需要移动剩下距离的一半，之后你需要移动当前剩余距离的一半。以此类推，你始终需要完成剩余距离的一半（见图14–8）。



图14–8

每完成一个阶段之后，还是会有剩余距离的一半是你没有完成的。那么这是否意味着你永远无法到达目的地？

数学家们总是喜欢在他们已经理解的观念的基础上再构建一个新观念，我们可以用接下来的方法将这种情况与无穷的自然数联系起来。我们可以将尚未完成的距离设为总距离的$\frac{1}{2}$，然后是$\frac{1}{4}$、$\frac{1}{8}$、$\frac{1}{16}$，以此类推直到“永远”。正如我们之前说的，自然数藏在永远之中。假设我们最终需要完成的距离是1公里，则我们整个行程的每个阶段如下所示：

	公里
阶段 1	$\frac{1}{2}$
阶段 2	$\frac{1}{4}$
阶段 3	$\frac{1}{8}$
……	
阶段 n	$\frac{1}{2^n}$
……	

我们有无穷个n，因此我们的行程也会有无穷个阶段。我们无法完全在列表中一一写出每一阶段的距离，但是我们在原理上把它们全写下来了，这也就是上面那个带有n的公式所表达的含义。然而，如果我们无法写出行程的每个阶段，我们还能完成行程的所有阶段吗？答案是，当然能，因为事实上我们最终会完成这个旅程。我们每天都会完成一些行程，可能是一些稍短的行程。比如，我不是每天都会离开家，但是我每个小时都会去冰箱很多次。

芝诺提出的另外一个相关的悖论，是一个关于阿喀琉斯和乌龟从A点到B点进行赛跑的问题。假定乌龟的起跑点在阿喀琉斯前方，比如说在点A_1，但是它跑起来很慢，因为它毕竟是一个乌龟。现在，阿喀琉斯必须首先到达

乌龟的起跑点。当他到达那里时，乌龟前进了一些，比如已经到达了点A_2。阿喀琉斯现在必须到达这个点。但是在他到达A_2的时间里，乌龟将继续前进一小段距离，比如到达了点A_3。接下来，阿喀琉斯必须到达A_3，但是在他这么做的时间里，乌龟又到达了A_4。在比赛过程中，每次阿喀琉斯到达我们刚刚确认的乌龟的位置时，乌龟就又前进了一些。那么，这难道就意味着乌龟一定会赢得比赛吗？

两个悖论都从一个看似合乎逻辑的观点，引发出了一个荒谬的结论。我们当然能开展一个旅程并抵达终点。如果乌塞恩·博尔特（Usain Bolt）和一只乌龟赛跑的话，显然是博尔特会赢。这些悖论的重点不是表明现实是错误的，而是表明论点中假定的逻辑是有错的。

这与希尔伯特旅馆问题是两种不同的类型。在希尔伯特旅馆中，旅馆可以是客满的同时仍能接纳一位新顾客。这个悖论听起来很荒谬，但这只是因为我们一开始对于无穷旅馆的直觉并不正确。

希尔伯特旅馆类型的悖论被称为证真式悖论，这意味着它由一个完全正确的论据产生出了一个看似矛盾实则正确的结论。而芝诺类型的悖论被称为证假式悖论，这里矛盾的结论是由一个看似正确但实际上并不正确的论据产生的。

这两种悖论的共同点是，当你开始考虑无穷时，奇怪的事情就会发生。希尔伯特旅馆发生在事物无穷大时，而芝诺悖论发生在事物无穷小时。在希尔伯特旅馆的例子中，我们很难让自己的头脑习惯事物无穷供应的想法，因为这在现实生活中是不可能发生的。无论鞋子、袜子、票据还是旅馆房间，它们在现实中都是有限的。然而在芝诺悖论中，我们认识到，如果我们允许一个漏洞存在，我们就有了无穷供应的事物，这个漏洞就是无穷小的事物。它们不能“是”无穷小，因为我们并不知道这意味着什么，但是它们可以“变得”无穷小。我们每天都在体验着这些无穷的事物，却从来没有意识到它们的存在。

无穷多的无穷小事物

在从A移动到B的悖论中，我们确实成功地从A抵达了B，这意味着我们确实覆盖了无穷数量的距离。但是其可行的原因是那些距离持续变得越来越小，因此我们在每段距离上花费的时间也持续变得越来越短。这是在现实生活中，而不是在想象的希尔伯特旅馆的世界中。在希尔伯特旅馆的世界里，我们不需要知道为什么我们有足够的时间来填满无穷数量的旅馆房间或者分发出无穷数量的票据，我们只需要知道我们能这么做就可以了。而在现实生活中，只有当我们花费在每件事上的时间无穷小的时候，我们才能每天做无穷数量的事。

例如，想象一下你正在从家去往火车站的路上，这段距离有1公里。我们假定你以每小时4公里的速度匀速前进，那么这段路程将花费你15分钟的时间。但是芝诺悖论会怎么说呢？

- 首先你必须先走完$\frac{1}{2}$公里，这将会花费你7.5分钟。

- 接下来，你必须走下一个$\frac{1}{4}$公里，这将会花费你3.75分钟。

- 然后，你必须走下一个$\frac{1}{8}$公里，这将会花费你1.874分钟。

- 再然后，你必须走下一个$\frac{1}{16}$公里，这将会花费你0.9375分钟。

- ……

你必须走完所有这些越来越小的距离，但是你走完这些距离所需的时间也越来越少。在你到达火车站前有多少段这种小距离呢？无穷多。如果我们在走完了有限多的小距离之后停下，实际上整段路程也只会剩下非常短的一小段距离了。

用这种方法来计算你需要花费多长时间才能抵达火车站显然是非常荒谬的，因为到了某一时刻，剩下的非常小的距离将比你的脚的尺寸还小。然而，它是一个非常重要的思想实验，引导我们得出这样一个认识：如果我们加和的事物正变得越来越小，那么将无穷多的事物相加并得到一个有限的结果是可能的。在现实生活中，我们无法发放无穷数量的票据，因为票据全是相同尺寸的。即使票据在变小，我们还是会花费独立的、离散的时间来发放每一张票据。因此发放无穷张票据这件事是我们无法完成的。我们无法吃无穷口巧克力蛋糕，因为即使我们把每口蛋糕的量变成无穷小，把蛋糕送到我们嘴里这段距离仍是不变的。（不过，或许我们也可以持续缩短盘子和嘴之间的距离，最终把我们的脸贴在盘子上。）

这里事实上包含两个难题。什么时候将无穷多的小事物相加是有意义的？如果这件事是有意义的，那我们怎么才能知道答案是什么呢？数学家们为此奋斗了数千年，最终在20世纪用微积分解决了这些问题。我们将在下一章讨论这些问题。

15　分裂

我曾经潜过几次水，非常享受鱼群在周围环绕的飘浮感。其间，我尝试了解了大海的自然运动，欣赏了珊瑚的构成。有时我会小心翼翼地在它周围游泳观赏，而珊瑚会像悬崖边缘的岩石一样突然掉落。我会由此产生一种奇怪的眩晕感，就像我站在一个悬崖边上真的要坠落了一样。我飘浮在海里，因而根本不存在坠落的危险，只是被掉落的珊瑚扭曲了感知。

数学中的某些概念也会像这样扭曲人们的感知，无穷就是其中的一个。一瞬间你认为你理解了正在发生的事情，下一个瞬间，你看待它的角度稍微发生一些改变，你就会觉得之前弄懂的一切全都消失了。当数学家们意识到他们脚下的基石可能不再存在的时候，他们也会体验到这样的瞬间。接下来，他们不得不赶快修复这些基石。思考无穷小的事物就会引发这种情况，因为数学家们可能会突然发现他们现在不能准确地阐述什么是实数了。

在上一章中，我们发现在有限的时间里我们可以做无穷多件无穷小的事情，而在有限的空间里我们也可以放下无穷多个无穷小的事物。实际上，一切事物都是由无穷多个无穷小的事物组成的。

对于我提出的这个论点，你是否认同？数学家们花费了很长时间来将这一切表达成一个满足数学逻辑标准的形式。这是一个令人满意的数学故事，许多不同的问题导向了同一个答案，或者说不同的线索指向了同一个嫌疑人。芝诺悖论引导我们思考将无穷多个无穷小的事物相加的问题。我们也会在其他地方遇到这个问题，比如通过将曲线分解成小的直线段或者将弯曲的形状分解成小的正方形来理解它们。

但是在本书的第一部分，我们忽略了另外一个问题，那就是无理数到底是什么。就像在通常的数学课堂中一样，我们只是简单地说它们是无限不循环的小数。事实上，这也是一个将无穷多个无穷小的事物加起来的问题。因为当我们给一个数字赋予越来越多的小数位数时，我们其实是在把越来越小的分数加和起来。

最初，我们需要无理数就是因为我们要填补所有有理数之间必然存在的“间隙”。结果证明，我们可以依靠这些有无穷位的小数来填补间隙。换言之，我们通过无穷加和无穷小的事物解决了这个问题。现在剩下的问题就是我们需要搞清楚这个无穷加和意味着什么。这个故事的奇怪之处在于，甚至连数学家们都没有完全意识到他们并没有解决这个问题，他们像我们一样就这么愉快地开始使用起了这些数字。数学家康托尔和戴德金

（R.Dedekind）都意识到，他们并不知道如何严谨地定义数字。戴德金看起来是在为数学课备课的过程中发现这个问题的。我明白那种感受。当你正准备教授某些事物的时候，你突然意识到有些事物你以为你是理解的，但实际上并没有充分理解到可以向其他人讲解它的程度。康托尔和戴德金面临的情况是，世界上还没有人充分地理解了它。幸运的是，他们有能力修复这些基石，因此整个数学世界得以幸存。不仅如此，他们对无穷小的事物的研究也使数学得以加速向前发展。

近似圆

当我的小侄子要开始上幼儿园时，他收到了一张推荐书单，上面写有一系列儿童数学读本。我和他一起读了玛丽莲·伯恩斯（Marilyn Burns）的《贪心的三角形》（*The Greedy Triangle*）。这本书的主角是一个三角形，但它厌烦了做一个三角形。于是它来到“变形器”面前，要求获得一个新的边。从此它变成了一个四边形。不可避免地，过了一段时间它又厌烦了做四边形。因此，它回去又要了一个边，于是它变成了一个五边形。接下来它又厌烦了，于是它又变成了六边形、七边形和八边形。

读这本书时我非常兴奋，因为我觉得我能预测到它会怎样发展下去。事实上，我的小侄子也变得非常兴奋。（也许是我的兴奋太有感染力？我希望是。）我几乎能看见他的小脑袋在高速运转，然后他大声叫道，“我知道将会发生什么！它会变成一个圆！”

但最终故事并没有按照我们想象的方向发展。我对此有一点儿难过。如果它最终变成了一个圆，这本书就可以被视为微积分的引言了。但实际上，更加有道德教育意义的事情发生了：它意识到过去的形态让它更加快乐，然后它就又变回了一个三角形。

我侄子的想法与数千年前巴比伦数学家试图算出圆形的面积时的想法是一样的。计算一个圆形的面积很难，因为它的轮廓是弯曲的。事实上，说明“一个圆形的面积”是什么意思都很困难。在某种程度上，这个面积的意思是“你能纳入其中的单位正方形的数量”。但是要做到这一点，你必须能够“融化”这些单位正方形，并将它们“倒入”一个圆形“模具”中。或者你可以反过来做：“融化”一个圆形，并将它“倒入”一个方形的“模具”中，观察这个方形占用了多大面积。这就是古人思考“求圆形面积”这个问题的方式。现实生活中偶尔也会出现这种情况。当我想要在一个方形烤盘里做蛋糕，而我找到的配方是圆形蛋糕的配方时，我就会遇到这个问题。我时常这么做，因为我有一个可调节的方形蛋糕盘，它能够制作从1吋到12吋的任意尺寸的蛋糕，但是我只有三个圆形烤盘。当然，我只需要使用一个圆形面积公式（而且我经常假定$\pi=3$）就能知道我该用多大的方形烤盘了。

但是最初人们是如何想出这个公式的呢?

一种不需要融化方形或者圆形，也不需要浪费蛋糕糊的方法就是，将圆形用有直线边的多边形来做近似。你使用的边越多，你近似的效果就越好。

例如，如果你使用一个与圆形内切或者外切的正方形来做近似，你会发现这些正方形对于这个圆的近似效果很差。换言之，如果你使用8吋圆形蛋糕的配方在8吋方形蛋糕盘中做蛋糕，你的蛋糕糊肯定会不够（见图15–1）。

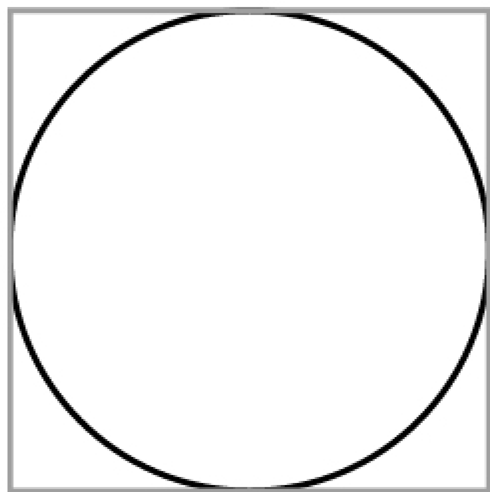


图15–1

我们可以算一算两者的面积到底差了多少。我们设定圆的半径为r。我们可以对内接正方形和外切正方形两种情况都做一下计算（见图15–2）。

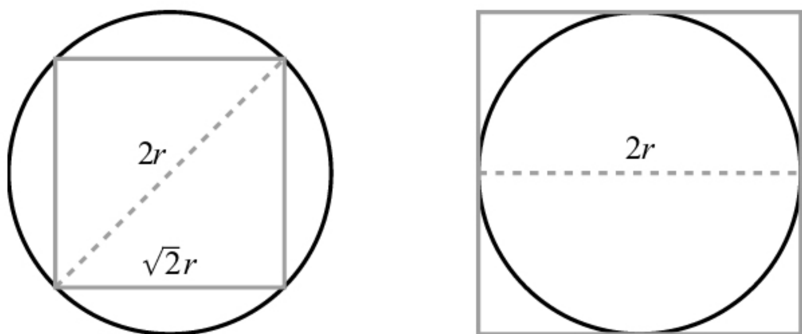


图15–2

圆形的面积是πr^2。内切正方形的面积是$2r^2$。外切正方形的面积是$4r^2$。

你可以利用勾股定理得到内接正方形的边长，从而计算出内接正方形的面积。或者你可以使用如图15–2所示带有虚线的三角形，将4个这样的三角形拼在一起得到一个虚线在外部围成一圈的正方形，这个正方形每个边的边长为2r，所以这个新正方形的面积就是$4r^2$。而我们所需要的内接正方形的面积就是这个虚线正方形的$\frac{1}{2}$。

这样我们就能得到圆形的“真实面积”和两个估算面积之间的对比。

$$\text{圆形} = \pi r^2 \approx 3.14r^2$$

$$\text{内接正方形} = 2r^2 \qquad \text{误差} \approx \quad 36\%$$

$$\text{外切正方形} = 4r^2 \qquad \text{误差} \approx +27\%$$

如果你使用一个正八边形来近似圆，其估算效果将会更好一些（见图15–

3）。

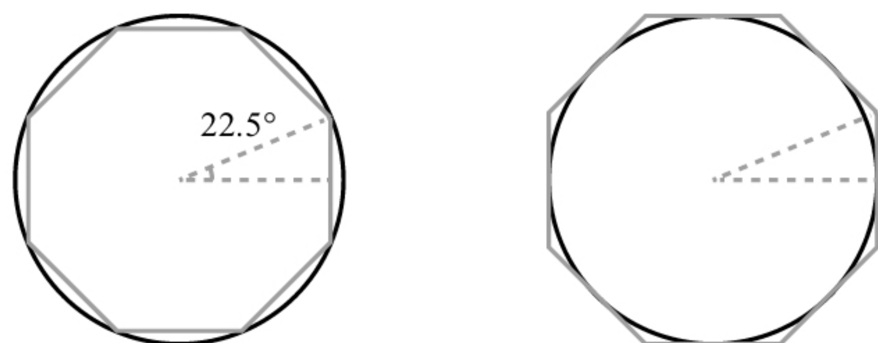


图15–3

在这种情况下，我们会得出下面的结果。

$$\text{圆形} = \pi r^2 \approx 3.14r^2$$

$$\text{内接八边形} = 2.83r^2 \qquad \text{误差} \approx 5\%$$

$$\text{外切八边形} = 3.31r^2 \qquad \text{误差} \approx +10\%$$

我们可以使用如图15–3所示的方法将这些八边形分割成16个直角三角形，从而计算其面积。圆中心的小角角度为：

$$\frac{360^\circ}{16} = 22.5^\circ$$

接下来，我们可以使用传统的三角学知识计算出三角形的边长。使用三角学可能会被认定为作弊，因为在人们考虑圆的面积的时代，三角学还没有出现。但是这里我们只是想证明一个观点，并不想模拟历史。

这里的基本理念就是，你使用具有越来越多的边的正多边形来模拟圆形，边越多，直线边就会越短，直线边与圆形的真实曲线边之间的误差就会越

小。圆的“真实面积”介于内接多边形的面积和外切多边形的面积之间。你可以通过加入更多的边来获得你想要的精度。

这也是计算一个由曲线构成的封闭空间面积的方法，或者说是计算曲线“之下”的面积的方法，如图15–4所示。

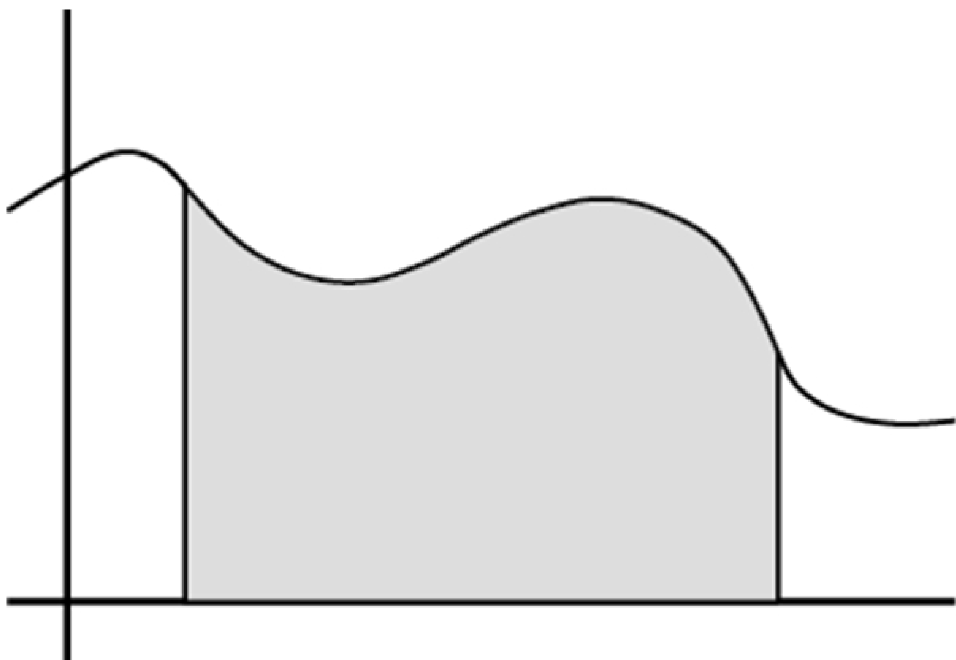


图15–4

具体做法是将曲线下的面积分割，通过边线为直线的碎片来近似这个面积，并且使用来近似曲线的那条直线边尽可能地短（见图15–5）。你分割的碎片越多，你近似的准确度也就越高。

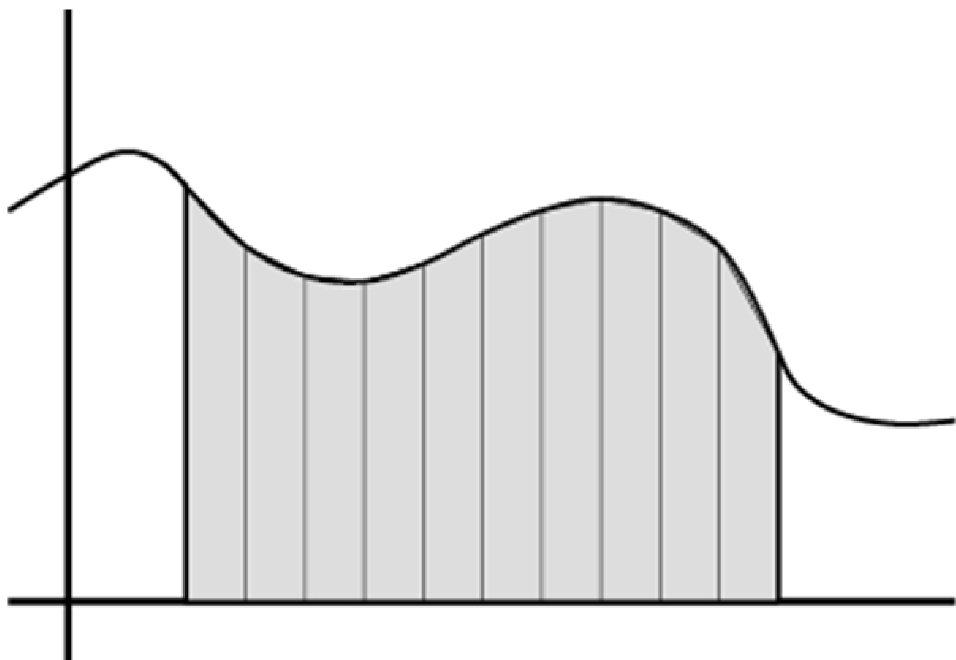


图15–5

但是你最终能得到精确的答案吗？你会越来越接近答案，但你永远不可能真正到达那里。因为直线边永远不可能精确地和曲线边匹配起来——真的是如此吗？

解决这个难题是驱动现代微积分发展的强大动力之一。起初答案看起来可能有点儿气人，但这就是展现抽象数学如何运作的典型案例。答案是，我们可以用适合我们的方法来定义“正确答案”是什么意思。这意味着至少它不应该造成任何矛盾，此外，它应该能够通过基本的行为测试，以使得它在最大限度上符合我们的直觉。例如，如果这条曲线其实是一条直线的话，那么我们通过上面那个方法计算出的答案就应该是精确的。

无论是圆、曲线的近似还是芝诺悖论，它们的关键在于，在所有的情况下我们都要一次又一次地重复一件事，每重复一次我们就前进一步，距离精确的答案也就更近一点儿。经过任何有限次的重复后，我们都不会得到最终答案，但是我们在持续地接近最终的答案。而在我们“完成了”这件需要永远进行下去的事情之后，我们需要为之找到一个意义。这里我们的直觉遇到了困难，因为我们不是永生的，因此我们也无法想象当我们到达最终的答案之后会发生什么。但是在这一章里我们会发现，如果确实存在一个答案是有意义的，我们不妨先认定它然后观察会发生什么——奇迹发生

了，这种做法填补了有理数之间所有令人烦恼的空隙！

有理数间的空隙

在第4章里我们曾经提到，如果我们只考虑由两个整数的比构成的数字，那么在它们之间总会存在空隙，即使这些空隙非常非常小，事实上，是无穷的小。这是什么意思呢？

首先，找到两个有理数之间的空隙并不难。当你画出一个边长为1（1厘米、1米、1千米……任何单位都可以）的正方形（见图15–6）时，你就能找到一个。

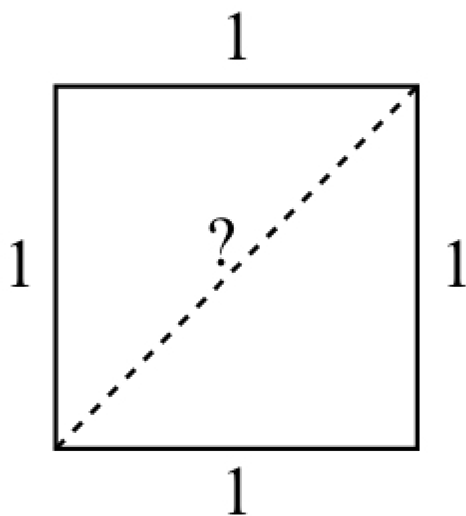


图15–6

这个正方形的对角线的长度是多少？如果我们在这个正方形里画出一条对角线，我们将会得到两个直角三角形。根据勾股定理我们知道，“斜边的平方等于其他两边的平方之和”。在这里，我们用d代表对角线的长度。由此可知：

$$d^2=1^2+1^2=2$$

所以，对角线的长度就是一个平方等于2的数，即它应该是2的平方根。但是我们发现，如果我们只考虑由两个整数的比构成的数字的话，这个数将

不会存在。下面是我们的证明过程。

证明：首先，我们假设与我们想要证明的论点相反的论点是正确的。因此，我们假设确实存在两个整数a和b，且$\sqrt{2}=\frac{a}{b}$。关键是，还要假设该分数是在它的最简形式下。换言之，你不能用分子和分母同时除以某个数来得到一个更简单的分数。

现在我们对两边取平方，得到：

$$2=\frac{a^2}{b^2}$$

$$\therefore 2b^2=a^2$$

到目前为止，我们还没有发现矛盾。现在我们知道，a2是某个数的两倍。这意味着它是一个偶数，同时a也必须是一个偶数。因为如果a是奇数，a2也将是奇数。

a是一个偶数意味着什么呢？这意味着它能够除以2，即$\frac{a}{2}$也是一个整数。我们设定：

$$\frac{a}{2}=c$$

$$\therefore a=2c$$

现在我们将它代入到上面的等式中，得到：

$$2b^2=（2c）^2$$

$$=4c^2$$

$$\therefore b^2=2c^2$$

现在，我们可以对b做我们对a做过的相同的推理。我们知道b^2是某数的两倍，因此它是偶数。这意味着b也必须是偶数。

现在我们发现a和b都是偶数。但是在最初我们就假设$\frac{a}{b}$是最简分数，这意味着a和b不可能同时为偶数。因此，证明与假设是矛盾的。

所以最初假设$\sqrt{2}=\frac{a}{b}$是错误的。这意味着$\sqrt{2}$不可能被写成一个分数，所以它是无理数。

因此，只是画了一个如图15–6所示的正方形，我们就不可避免地遇上了边长为无理数的情况。这表明，在有理数之间肯定存在至少一个小空隙，而2的平方根就在这个小空隙里。如果我们画一个半径为1的圆，我们会碰上另一个有理数间的空隙。因为我们所画的线的长度（也就是圆的周长）将是无理数。虽然这个证明过程比刚刚证明正方形的对角线长度是无理数更难一些。

描述一个圆的周长有一点儿困难，但是一个正方形的对角线长度并不存在什么争议。如果你想象一个正方形的话，它的对角线肯定有一定的长度。如果这个长度不是一个数字的话，那么我们的数字体系肯定是不完善的。这些空隙还会引起其他的问题。一个奇怪的例子是，它们意味着你有可能画出两条互相交叉但是并不相交的线，因为本该相交的地方可能恰好在一个小空隙里。它还关系到我最喜欢的一个数学应用实例：迷你胡萝卜。

迷你胡萝卜存在吗？

我曾经在一个数学会议上大嚼迷你胡萝卜。为了避免脸上沾上太多的饼干屑，我专门买了这些迷你胡萝卜在茶歇时吃。结果我和一些人争论起了迷你胡萝卜到底是真的还是假的。那时我还没有移居美国，也还不知道原来在美国你可以买到人工切割的迷你胡萝卜。他们把普通的胡萝卜按迷你胡萝卜的尺寸切割，形状是标准的圆柱形，尾部规则又饱满。在美国，人工切割的迷你胡萝卜比英国版的迷你胡萝卜常见得多。英国版的迷你胡萝卜是还没有长到很大的胡萝卜。这种迷你胡萝卜常常是完整的，有表皮而且尾部仍然连着绿色的茎叶，所以看起来就是整个儿的胡萝卜。小时候，我们在花园里种了一些胡萝卜。当我们把它们拔出来的时候，有些胡萝卜确实会比正常的胡萝卜要小得多。

对于我来说，这些关于胡萝卜的看法都是显而易见的事实。但是我的美国同事对于这些事实感到很惊讶。不过因为我们都是数学家，所以我很容易就能用中间值定理来说服他。基本上是这样的：胡萝卜都是从什么都没有连续生长起来的，所以在某一时刻它们肯定是很小的。

当然，因为我们都是数学家，所以很快就出现了反驳的声音。他们问我，我是怎么知道胡萝卜是连续生长的？我怎么知道它们不是由绿色的胶质组成，直到到达某一个特定的尺寸，才自发地在那一时刻变成了一个胡萝卜？这有点儿像蝴蝶破蛹而出后就呈现出完全态了一样。数学的论证往往来自数学家和质疑者或者数学家和自作聪明的人之间的有来有往的辩论。我们都知道这整个争论实际上就是一个数学上的笑话，但是这仍然是我最喜欢的数学原理应用实例。（从这一点你可以准确地推断出，我并不是一个会被伟大的原理应用而激励的人。）

那么，中间值定理到底说的是什么呢？下面我会用另外一种更加数学化的方式来表达中间值定理。如果我们画出$y=x^2$的图像，那么它看起来会如图15–7所示的那样。这个形状叫作抛物线。

如果我们画出$y=2$的图像，那么它看起来会如图15–8所示的那样。

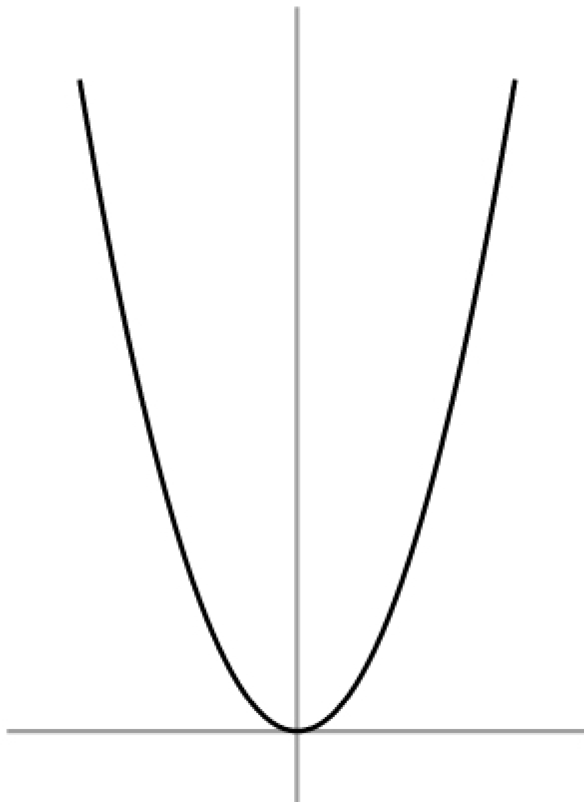


图15–7

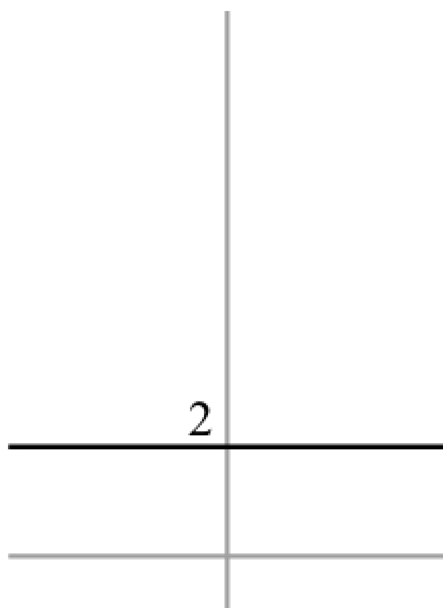


图15–8

现在我们将它们叠加在一起，并且观察它们相交的地方（见图15–9）。

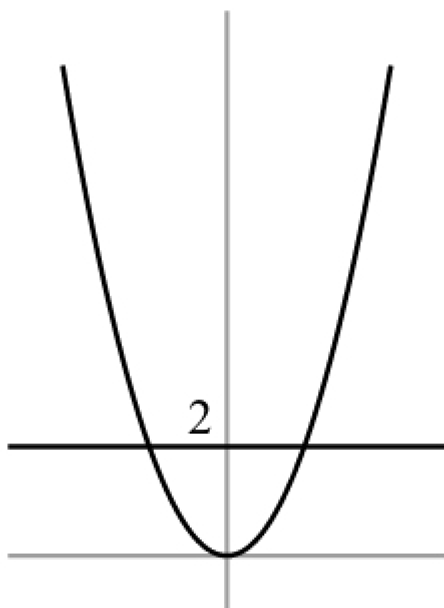


图15–9

如果我们只有有理数的话，我们知道没有能够满足$x^2=2$的有理数x。这意味着图中两条线在任何位置都不会相交。抛物线上会有一个小空隙，因此直线不会在任何位置接触到抛物线。这很诡异：直线在不碰到抛物线的情况下从一侧到了另一侧。这让我想起了纽约的地铁。铁路线在各种各样的地方相互交叉，但是乘客并不能在交叉处换乘。而在伦敦的地铁里，只要铁路线相互交叉，大多数情况下人们都能在交叉站换乘地铁（至少伦敦市中心没有任何例外的情况）。

我刚刚说过，空隙在抛物线上，那么它有可能在直线上吗？实际上，两条线上都有空隙，而且两条线上的空隙也不仅仅是交叉处那一两个点。空隙无处不在，因此有各种各样的直线与抛物线交叉但不会相交的情况，将直线从2移动到3你就会发现另一个空隙。这就是我们想要填补那些空隙的原因之一。我们想要达到的效果是，如果我们画出两条线而且它们交叉，那么它们就必然在某个地方相交。这就是中间值定理。这么命名是因为其最根本的形式是，如果你从高度A持续增长到高度B，那么你一定经历过A和B之间的任何一个高度。任何一个目前身高六英尺的人，在过去的某一个时刻肯定曾经身高五英尺。胡萝卜肯定在过去的某个点是迷你胡萝卜。

空隙无处不在

现在我们知道，在有理数之间至少有一个“空隙”，$\sqrt{2}$的值就属于这种空隙之一。我们在第4章已经提到过，在任意两个有理数之间都有一个无理数。这意味着无理数必然是无处不在的，无理数将这种“空隙”插进所有的有理数之间。实数在某种程度上可以说是条纹状的——有理数和无理数相间出现。当然，我们必须小心对待这句话的真正含义。

如果一些自作聪明的人抛给我们两个有理数，我们肯定能够找到一个介于它们之间的无理数。无论这两个有理数有多接近彼此。我们下面就将展示这个过程。假设自作聪明的人抛给我们的两个有理数为a和b。我们所需做的就是在a上加一个很小的无理数，因为一个有理数加一个无理数还是无理数（见图15–10）。

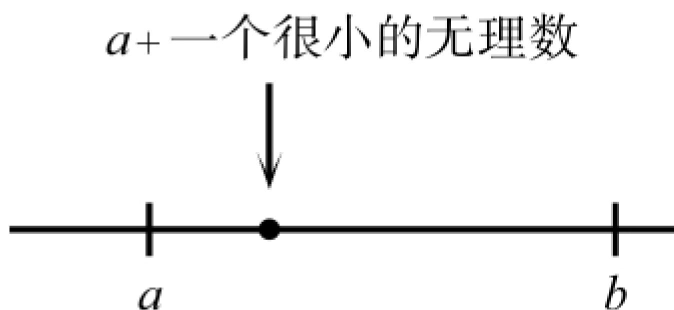


图15–10

这个数需要有多小才能让a加上它得到的数落到空隙里呢？它只需要比b–a小就可以。这很简单。我们可以通过用$\sqrt{2}$除以一些大数得到我们想要的无理数。比如，我们可以构建$\frac{\sqrt{2}}{100}$、$\frac{\sqrt{2}}{1000000}$和$\frac{\sqrt{2}}{10000000000}$。我们使用的除数可以是任意大的数，因此我们得到的无理数可以是任意小的数。无论我们那些自作聪明的邪恶对手把a和b设计得多接近，我们都能够用这种办法应对。

无理数除以有理数仍然是无理数。这和一个无理数与一个有理数相加仍是无理数的事实类似，这个论述也可以用反证法来证明。

这里我们想说的是，我们可以找到一个无穷接近a的无理数。如果“无穷小”是一个有效的量，我们可以称其为ε（希腊字母，epsilon）。这样，我们的问题就变成了，在距离aε远的范围内找到一个无理数。需要注意的是，这个ε仅在它作为一个概念时是有效的，它不能作为一个量来使用。所以我们使它逻辑有效的方式就是将ε转化成我们自作聪明的邪恶对手赋予ε的任何值——无论他们选择的值有多小。他们越想要挑衅你，挑选的值就会越小。接下来，你的任务就是表明无论他们的ε有多小，你都能在距离aε远的范围内找到一个无理数。如此，我们就找到了一个万无一失的制胜策略，无论那些自作聪明的人用多大的a和b来挑战我们，我们都知道如何在它们之间找到一个无理数。这就是现代微积分高度缜密性背后的主要支撑，使其无论处理无穷大的事物还是处理无穷小的事物都能够适用。无穷大实际上意味着“大到我们不会被自作聪明的人战胜”，同样地，无穷小实际上意味着“小到我们不会被自作聪明的人难住”。

根据数学传统，自作聪明的人用来为难我们的小距离常常被称为ε。这种情况非常多，所以“ε证明”成为这类证明的一个代号。ε代表自作聪明的人抛向我们的一个任意小的数字。毕竟向我们抛出一个大数字是一个失败的策略，因为我们在这种情况下更容易取胜。这个做法可以追溯到数学家们试图使用“无穷小的小量”来精确描述这类观点。然而，ε只是一个尽可能小的数字，而不是一个确定的小数字。因此，ε有时也被通俗地称为“任意小量”。数学家们会说这样的话，比如“几个月来我每天都感觉我距离完成这篇论文只差ε了”。事实上，有一个数学的笑话是下面这样的，“设ε为一个大负数”。如果你已经累死累活地花费了好几年时间来做ε证明，那么这个假定足以引起一阵歇斯底里的恐慌。

这种类型的论证与解方程或做几何题有很大的不同，它感觉上就像在绕开问题或者说从来没有真正解决问题。这是非常真实的感受，当学生们第一次遇见微积分时他们常常感到非常不安。因为这不是他们预期中的数学的样子。你在第一次遇见微积分的时候，它通常表现为一套条理清晰的规则和步骤，用来计算诸如图形的坡度和曲线下面积这样的事物。但为什么这样做能得到正确答案？这个问题被束之高阁，因为它太难以想象了。难以想象的原因是你必须将你的思维集中到无穷小这个概念上。我们已经了解到无穷大的事物是如何让我们感到困惑的，现在是时候了解那些无穷小的事物了。

想象一下，你被要求完成一个远距离射击靶子的测试。最初的时候你被允许有一些失误，但是为了通过测试，在某一个特定的时间点后，你必须能够保持无穷次地命中目标。如果你通过了这个测试，靶子就会变小，然后你需要再次挑战。如果你再次通过测试，那么靶子就会变得更小。靶子会持续变得越来越小，而测试也越来越难。你被认定具有击中目标的能力的

唯一标准就是无论靶子变得多么小，你都能通过这个艰难的考验。

射击训练

我们用某个会变得无穷小的事物来模拟这个射击挑战。比如$\frac{1}{n}$，当n越来越大时，$\frac{1}{n}$就会越来越小直至无穷小。我们发现这与“除以无穷，得到零”很像，而现在我们可以给这句话赋予一些合理性了。根本观点在于$\frac{1}{n}$永远不会真正等于零。如果0是我们射击靶的靶心，无论我们邪恶的对手把$\frac{1}{n}$变得有多小，我们仍旧有办法保持击中靶子。仅仅击中靶子一次是不够的，我们要做的是允许在最开始有一些失误的情况下，做到持续无穷次击中靶子。你可能想知道多少次失误算是“一些”。答案是，任意有限的数字。因为当我们谈论“永远”的时候，一开始的有限次的失误相比之下就显得无足轻重了。

例如，假设邪恶的对手将射击靶的半径设置为只有0.1。那么我们需要击中的范围就如图15–11所示：

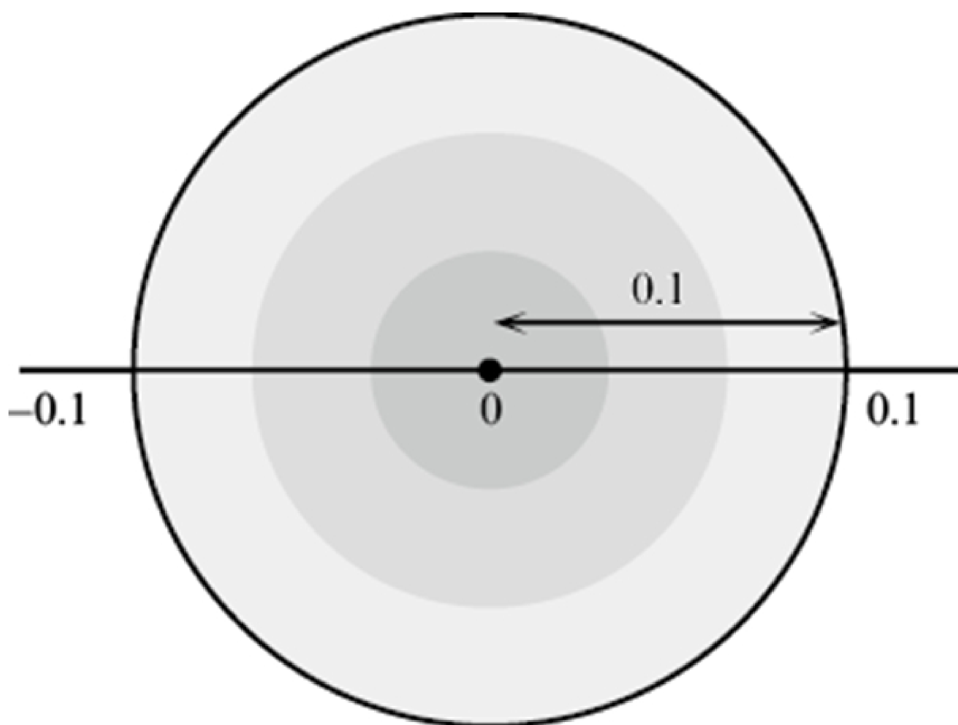


图15–11

我们的靶子应该只是实数轴的一部分，但是我将它画成一个圆形靶，因为这样看起来更像是在做射击练习。

我们进行射击的尝试由逐渐增大的n构成，就像下面这样：

$$1、\frac{1}{2}、\frac{1}{3}、\frac{1}{4}、\frac{1}{5}、\frac{1}{6}、\frac{1}{7}、\frac{1}{8}、\frac{1}{9}\ldots$$

我们第10次尝试的射击靶半径是$\frac{1}{10}$，也就是0.1。如果我们的射击成绩是这个数字，那代表着我们的子弹会擦过射击靶边缘。我们首次明确击中靶子的尝试是下一次，也就是$\frac{1}{11}$。从这里开始我们的成绩就很好了。接下来的尝试都会比这个数小，所以它们都会命中射击靶。

我们通过了这次考验。但是邪恶的对手是顽强的。他们用一个新的更小的

射击靶替换了旧的，新靶只有0.001宽。这次我们将花费更长时间来击中射击靶，但这是被允许的。记住，我们可以有任何有限次地失误。现在，会擦过射击靶边缘的尝试是$\frac{1}{1000}$，这和射击靶的半径0.001刚好一样。而下一次的尝试是更小的$\frac{1}{1001}$，这次我们就会击中射击靶了。接下来的每次尝试都会更小，所以都会击中射击靶。结果我们又通过了考验。

现在看起来，这个过程将会无休无止地进行下去。但是我们可以提出一个聪明的机制，这样无论邪恶的对手让目标变得多么小我们都仍能取胜。这有点儿像我们曾经证明的，无论a和b之间的空隙有多小，在它们之间肯定存在一个无理数。这次我们想要证明，无论ε有多小，我们都能够通过尺寸为ε的射击靶的考验。再说一次，ε不是一个固定的无穷小的小尺寸，因为“固定的无穷小的小尺寸”这句话本身就是不准确的。它是一个尽可能小的尺寸，小到自作聪明的邪恶对手想要的程度。

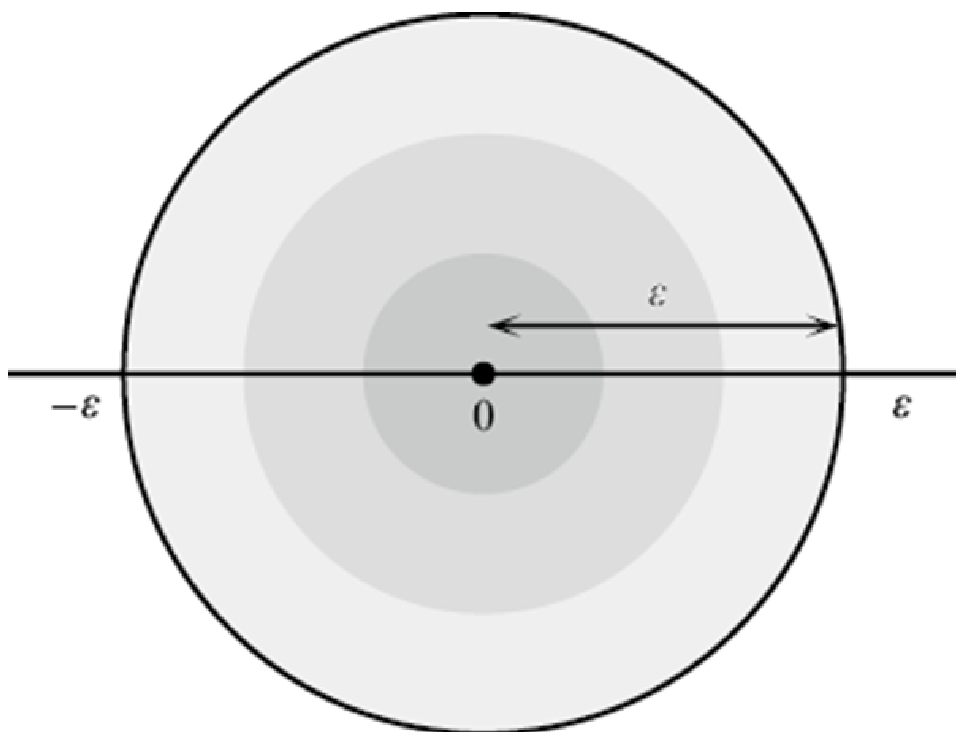


图15–12

这也是一个寻找时间点的问题，越过这个时间点，我们所有的$\frac{1}{n}$都会击中射击靶。现在，因为$\frac{1}{n}$会随着我们一次次尝试变得越来越小，我们只需要找到任意一个能击中射击靶的点。因为我们知道从这个点开始，我们在后面的尝试里就总能击中靶子了。我们能否找到第一个击中的点不重要，我们只需找到任意一个击中的点就可以了。

换言之，我们必须找到满足$\frac{1}{n}<\varepsilon$的n。我们当然可以令n非常大，大到我们需要的程度。这就像我们通过用$\sqrt{2}$除以一个足够大的数来得到一个小的无理数一样。

如果我们想要更精确一些的话，我们可以算出需要除多大的数字才能得到一个足够小的数。在这个问题里我们正试图找到$\frac{1}{n}<\varepsilon$。如果把它变换一下，这个不等式就会变成$\frac{1}{\varepsilon}<n$。所以我们只要找到一个比$\frac{1}{\varepsilon}$大的n就可以了，无论这个数是多少。我们知道我们总能做到。因为n是越变越大的，而且永远不会终止。

这意味着在原理上我们知道，我们总能通过射击考验，无论邪恶的对手们选择的射击靶有多小。

在数学上，我们说0是下面这个数列的“极限”：

$$1、\frac{1}{2}、\frac{1}{3}、\frac{1}{4}、\frac{1}{5}、\frac{1}{6}、\frac{1}{7}\ldots$$

这并不意味着这个数列中的数字最终会变成0，但是它意味着数列中的数字会越来越接近0。这和我们刚刚描述的射击挑战具有完全相同的含义。

再谈一次蛋糕

我们首先吃一块蛋糕的$\frac{1}{2}$，然后吃剩下的$\frac{1}{2}$（全部蛋糕的$\frac{1}{4}$），然后吃新剩下

的$\frac{1}{2}$，然后再吃新剩下的$\frac{1}{2}$……以此类推。通过这种方式，我们是否能使蛋糕永远吃不完？现在我们可以用前面描述的方法赋予这个问题一个严格的答案了。也许你会认为你的蛋糕能永远吃不完，但是剩下的蛋糕将会越来越小以至于可以忽略不计。因此从根本上讲你将会吃下整个蛋糕。

这个过程可以用一个令人满意的方式在方形蛋糕上画出来（见图15–13）。只要你不介意有些部分是正方形的，而另外一些部分是长方形的。

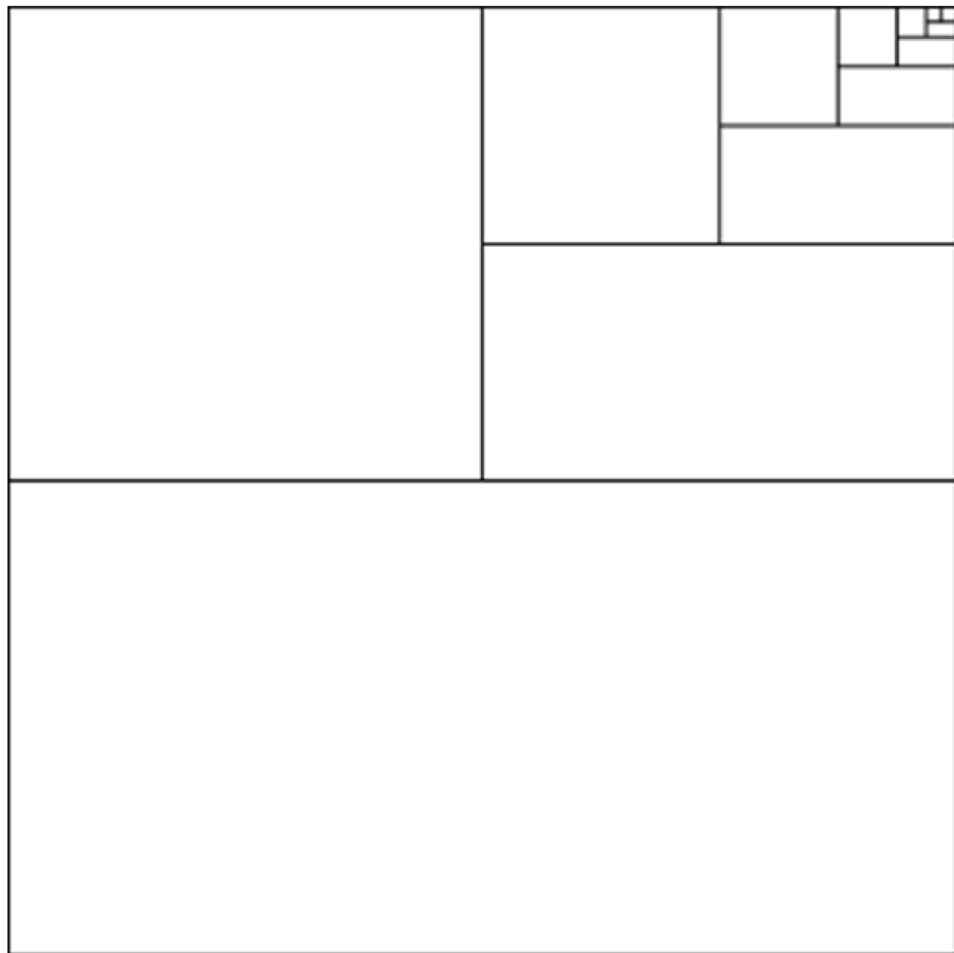


图15–13

我们的问题是：在我们到达永远后我们一共吃下了多少蛋糕？随时间推

移，我们在每个阶段依次吃下了这些量：

$$\frac{1}{2}、\frac{1}{4}、\frac{1}{8}\ldots$$

但是这次我们关心的并不是每个阶段我们吃下的量，而是我们吃下的总量。每个阶段后，我们吃下的总量为：

$$\frac{1}{2}$$

$$\frac{1}{2}+\frac{1}{4}$$

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$$

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}$$



现在我们要宣布，如果我们“以此类推直到永远”，我们将吃下所有的蛋糕。在这里我们的意思是，我们将越来越接近吃下整个蛋糕。根据之前射击挑战的方法，我们可以说这里的“接近”是接近到任何自作聪明的人无法质疑的程度。从数学上讲，我们表达的是我们吃下的蛋糕的量的极限是1，也就是整个蛋糕。

自作聪明的人再次来挑战我们了。这次，射击靶靶心位于1的位置。他们在最开始将射击靶的半径设为0.1。

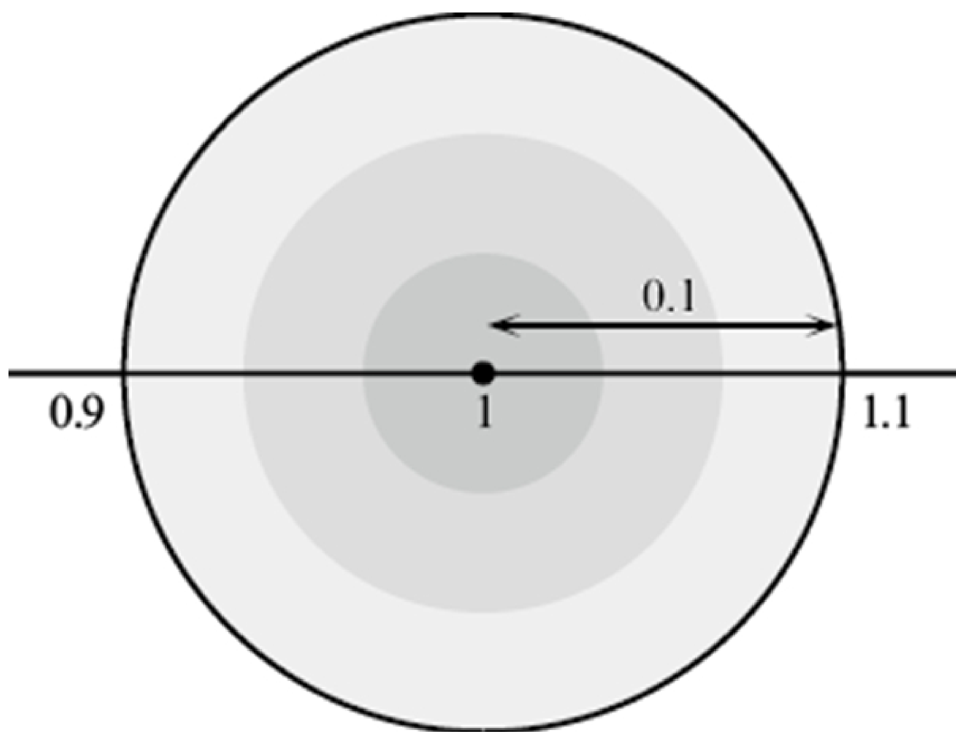


图15–14

我们需要证明，在被允许的有限次失误后，我们能持续击中这个射击靶。这次，我们的尝试将按照前面每次吃下的蛋糕总量的列表逐行进行。你可能会想，我们需要把那些分数都加起来得出每个阶段的答案才行。但是我们可以更懒一些（数学的关键就是尽可能地懒）。考虑还剩下多少蛋糕会更容易一些，而且这个方法能够告诉我们每个阶段我们距离靶心还有多远。

在第一次尝试之后，蛋糕还剩下$\frac{1}{2}$，所以我们距离靶心还有$\frac{1}{2}$的距离。鉴于射击靶的半径只有0.1宽，$\frac{1}{2}=0.5>0.1$，这说明这次尝试没有击中射击靶。

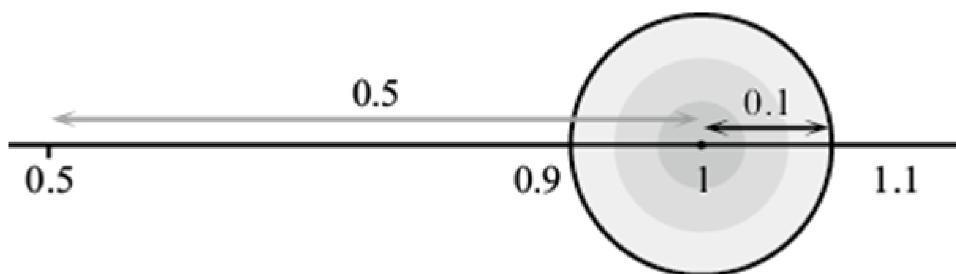


图15–15

我们的下一次尝试会使我们更接近些，这次我们还剩下蛋糕的$\frac{1}{4}$，这表明我们距离靶心1还有$\frac{1}{4}$＝0.25的距离。0.25仍旧比0.1大，所以我们还是没有击中射击靶。

再下一次尝试后，我们距离靶心还有0.125，仍然没有击中射击靶。再下一次距离靶心还有0.0625——终于比0.1小了，因此这次射击距离靶心1足够近，落在了射击靶上。在接下来的每一次尝试中，我们都会更接近圆心，因此我们将永远击中射击靶。所以我们通过了这次考验（具体过程见图15–16）。

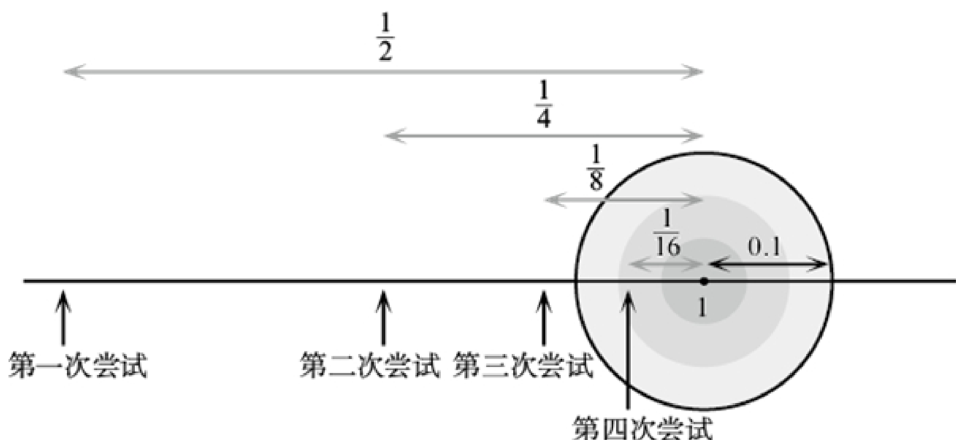


图15–16

事实上，无论给我们的射击靶是什么尺寸，我们总能通过考验。现在，假设自作聪明的人选择了宽为ε的射击靶。这就有一点儿困难了，因为我们

需要制定一个制胜策略来保证无论ε是什么我们都能获胜。当射击靶的尺寸是一个特定值的时候，我们只要一直试就能发现每个阶段剩余多少蛋糕，或者距离靶心还有多少距离，并在找到小于射击靶半径的那个数时停止。比如前面我们尝试了射击靶半径为0.1的情况。现在，我们需要制定一个适用于任意一个ε的策略。为此，我们需要在不知道ε是什么的情况下，知道什么时候剩下的蛋糕比ε小。这就是数学开始让你头晕的地方，因为我们的研究对象不再是一个准确的数字，而是变成了一个字母，而你甚至并不知道这个字母是什么。

我们意识到在每一次吃过蛋糕之后，剩下的蛋糕量都发生了减半。因此蛋糕剩下的量的变化如下所示：

$$\frac{1}{2}、\left(\frac{1}{2}\right)^2、\left(\frac{1}{2}\right)^3、\left(\frac{1}{2}\right)^4\ldots$$

因此在第n次吃过蛋糕之后，剩余的量为：

$$\left(\frac{1}{2}\right)^n=\frac{1}{2^n}$$

我们现在可以非正式地说，“随着n越来越大，2^n也随之变得‘巨大’，因此通过让n变得更大，我们总是可以使剩余量比任何ε都小”。

为了更加精确，我们需要获取一个点，此点的蛋糕剩余量小于ε。用数学式来表达，即：

$$\frac{1}{2^n}<\varepsilon$$

也就是：

$$\frac{1}{\varepsilon}<2^n$$

我们知道，只要n比2大，2^n总是比n大。因此如果我们让n满足$n>\frac{1}{\varepsilon}$，接下来我们就能得到这串不等式：

$$2^n > n > \frac{1}{\varepsilon}$$

这个不等式给出了我们需要的结果。我们没能找到击中射击靶的第一时刻，就这一方面而言，这并不是一个最有效的途径。但是这并不重要。我们只需要知道现在我们能够永远击中目标，而不需要确切地知道这是从什么时候开始发生的。对于这个结果，如果你喜欢偷懒就会感到满意，如果你喜欢精确就会感到不满。我两者都喜欢，但是偷懒取胜了。

由此，我们找到了一个无懈可击的策略。使用这个策略，我们可以通过任何尺寸的射击靶的考验。这意味着，在数学的定义中，剩余蛋糕量的极限为0，因此我们真正吃下的量的极限为1。这就是我们赋予“如果我们坚持这样做直到永远会发生什么事”的意义。

注意，在这里，“极限”并不意味着边界。换句话说，这并不意味着这是我们能吃的最大的蛋糕。我们在这里真正吃下的蛋糕的量的极限是1，仅仅是因为一开始我们就只有一个蛋糕。使用“极限”这个词是为了表达某件事情会一直发生直至到达某个点就不再变化了。

运用这个解释，我们同样可以解决芝诺悖论中从A移动到B的难题。在芝诺悖论中，我们必须走过全部距离的一半，然后是剩下距离的一半，以此类推。关键在于弄清我们在旅行的每个阶段所花费的时间。如果我们在整个过程中保持相同的速度，那么每个阶段所花的时间都将是上一阶段的一半。因此即使我们必须经历无穷个阶段，但由于每个阶段上所花的时间会变得越来越短，所以最终我们做这件事的总时长还是有限的。事实上，这与吃巧克力蛋糕的例子基本上是相同的。假设第一阶段会花费半小时，第二阶段会花费$\frac{1}{4}$小时，第三阶段花费$\frac{1}{8}$小时（虽然我们并不经常用八等分来测量时间），以此类推。如果我们将这些时间无穷次相加，我们就遇到了“自作聪明的射击挑战”的情形。这个问题的答案与吃巧克力蛋糕的答案相似。虽然要经过无穷个阶段，但是所有的时间相加起来为1小时。

因此，我们发现我们可以在有限的时间里做无穷的事，只要我们每天从一个地方去往另一个地方就可以了，除非我们整天都躺在床上。但是这就与我们讨论的问题无关了。

循环小数

现在，我们可以将射击挑战问题的解决方法用到循环小数上。我们来考虑一个最容易造成混淆的数0.99999999…，这个数也写作0.9。也许很多时

候人们会告诉你，这个数“等于1”。这句话的真正含义是什么呢？我们一般将这个循环小数看作“0.999…，其中9无限循环”。现在我们已经知道该如何用缜密的数学方式来处理这个问题了。这就像吃蛋糕一样。我们不断增加小数的位数。增加第一个小数位后我们得到了0.9，增加第二个小数位后我们得到了0.99，增加第三个小数位后我们得到了0.999，以此类推。$0.\dot{9}$ 的定义就是这个过程的“极限”。因此，当我们说 $0.\dot{9}=1$时，我们所说的就是这个过程的极限是1。这并不意味着我们确实能够到达1。这意味着，如果邪恶的对手出现了，并为我们设立了一个靶心在1这个点的射击靶，那么无论他们设定的射击靶有多小，我们都能够通过射击挑战。

我们来试试这种情况。假设第一个射击靶宽为0.01。于是，我们的射击靶看起来将如图15–17所示。

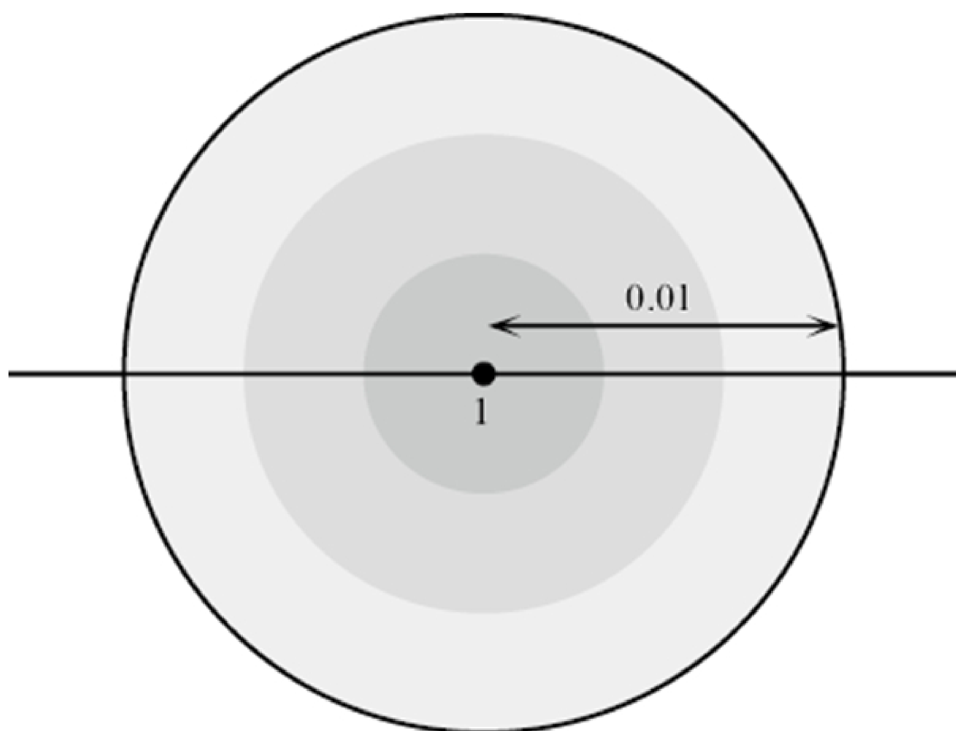


图15–17

我们第一次尝试的是0.9。它距离圆心1有多远呢？距离为0.1，这比0.01要大。因此我们没射中射击靶（见图15–18）。

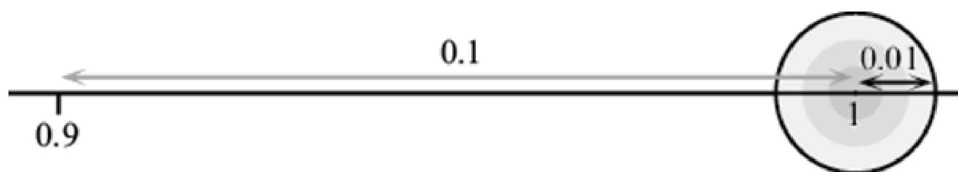


图15–18

我们的下一次尝试是0.99。这次离圆心1的距离为0.01，这意味着我们的子弹恰好擦过了射击靶的边缘（见图15–19）。

图15–19

我们的再下一次尝试是0.999。这次我们离圆心1的距离只有0.001，这意味着我们这次无疑击中了射击靶（见图15–20）。此外，我们所有后续的尝试均会更加接近圆心，因此我们知道，从现在起我们未来的所有尝试都会击中射击靶，也就是说，我们通过了射击挑战。

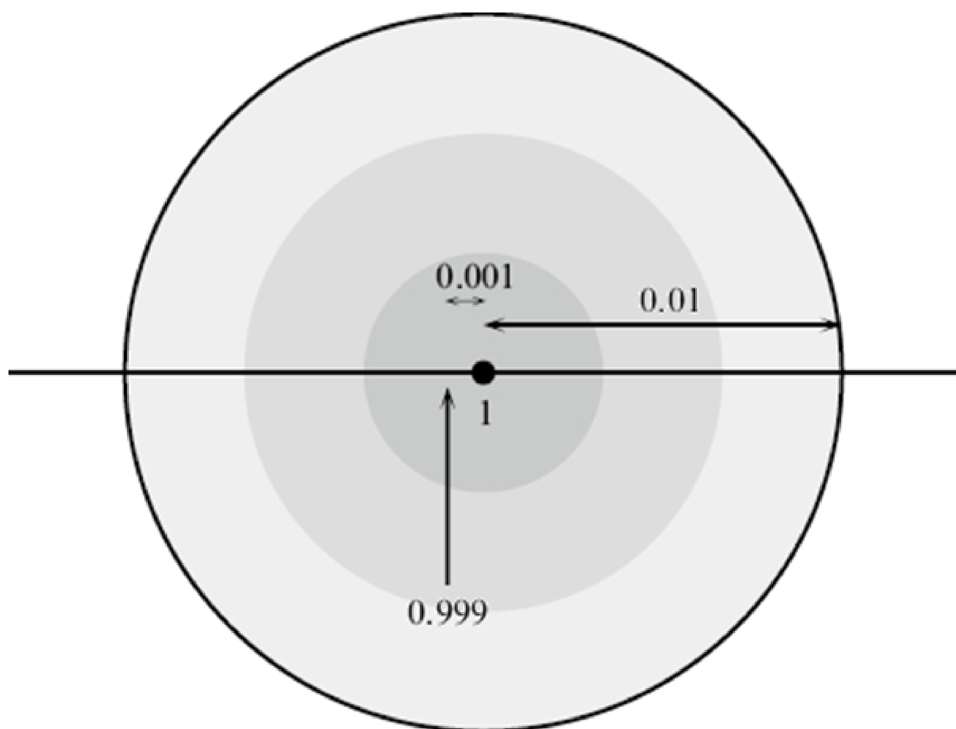


图15–20

现在，我们需要找到一个策略来保证无论射击靶有多小我们都能通过这项考验。我们假设邪恶的对手设定的射击靶半径为ε（见图15–21）。那么我们从什么时候起就能一直击中射击靶呢？

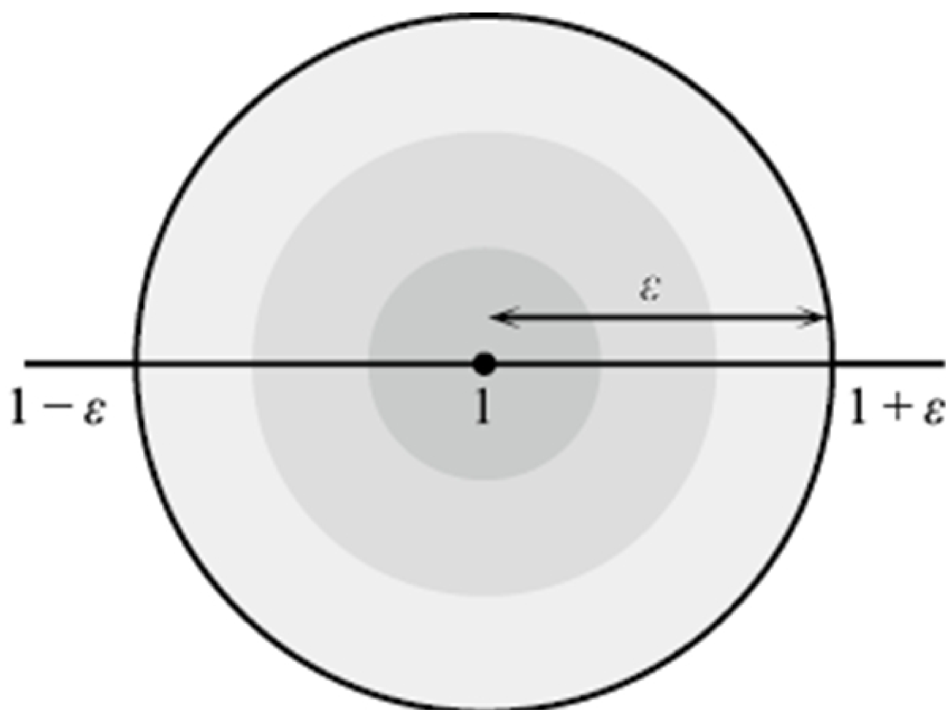


图15–21

我们知道，一旦我们的误差范围小于ε，我们就能击中射击靶。也就是说，我们需要找到一定数量的0，让0.0000…001＜ε。这是确定可行的，因为我们可以根据自己的需要设定0的相应数量。

这与蛋糕的例子非常相似。这里我们不再每次对剩余部分取半，而是用剩余部分除以10。因此我们需要n满足下面的条件：

$$\frac{1}{10^n} < \varepsilon$$

相当于：

$$\frac{1}{\varepsilon} < 10^n$$

与上次相似，我们利用$10^n > n$这个事实，如果我们让n满足$n > \frac{1}{\varepsilon}$，那么：

$$10^n > n > \frac{1}{\varepsilon}$$

这个n就是我们想要的答案。

如果你认为0.999…永远也不会“真正”到达1，那么你是正确的，但是你没有理解这个问题的关键。实际上这个循环小数的定义里已经包含了这个事实：这个循环小数的极限等于1并不需要它本身真正“到达1”。

另一个考虑循环小数的方式是问下面这个问题：如果0.999…不是1，那么它是什么？它不可能小于1，因为我们已经证明0.999…与1之间的距离可以比任何一个可能的空隙尺寸都小。所以无论你认为0.999…和1之间的空隙有多小，实际的空隙都不会比0大。那么我们现在有两个选择：

- 宣布0.999…没有意义，而且它并不是一个真正的数字。

- 宣布0.999…就是1。

我们来比较一下这两个选项。如果我们选择了第一个选项，我们也将无法解释任何其他的无限小数，那么我们将无法填充有理数之间的空隙。如果我们选择了第二个选项，我们将顺利填补有理数之间的空隙，而且根本不会发生任何逻辑上的错误。我们只是感觉有一点儿别扭。如果你还记得0.999…只是“n趋于无穷时，具有n位小数的0.99…9的极限”的简写，那么可能你的别扭感能稍微减轻一些。这样，你就可以开始享受这个旅程了。

如果我们使用二进制小数取代十进制小数，那么蛋糕的例子就变成了一个循环小数的问题。这是因为$\frac{1}{2}$在二进制中是0.1，$\frac{1}{2}$再加上$\frac{1}{4}$将得到0.11，再加上$\frac{1}{8}$将得到0.111，以此类推。如果我们坚持一直这样吃这个蛋糕，那么我们吃下的蛋糕量就是循环二进制小数0.1111…。根据我们上面的讨论，我们能够证明这个数的极限为1。

撇开“永远”

我们尝试想象了当某些事物持续进行到永远时会发生什么。对此，我们已经进行了很多练习，包括发放票据、为新到的客人在希尔伯特的无穷摩天大楼里安排房间等。想象这些事情是一个很难的锻炼，因为这些涉及永远

的事情没有一个会在现实中发生。我们要如何想象当一些不可能的事变成可能时会发生什么呢？

小说中有一个有趣的隐含机制，它能改变现实生活中的某个事物，同时保持其他所有事物都不变。例如，超人来到地球，但是地球上的所有其他人都还是正常的人类。还有，某人建造了一个时间旅行机器，但是所有其他人仍过着正常的生活。我们在关于“永远”的思维实验里需要做的正是这类特殊的事。

无论是从数学的角度看，还是从逻辑的角度看，这都是非常困难的。因为在逻辑上讲，如果你将一些错误的事变成正确的，那么所有事都会变成正确的。这有点儿像，如果你令$1=0$，那么所有数就都等于0了。

在一场争论中，某些人可能会说“如果你是对的，那我就是希巴女王”。他们说的实际上是，如果你是正确的，那么所有其他事也就都变成正确的了，包括像我是希巴女王这样完全不切实际的事。

在逻辑中，你不能只是选择某个错误的事物，然后强行令其正确，并且认为这不会产生任何后果——它的后果是你的逻辑体系将不再自洽。在维持逻辑体系自洽的情况下，让错误的事情变成正确的事情的唯一途径就是令所有事物都变成正确的。此时“正确”和“错误”将具有相同的含义，而数学和逻辑的世界也就崩溃了。这就是为什么想象希尔伯特旅馆和无穷的饼干都只是思维实验，而不是逻辑观点。这也是为什么它们可供争论，以及为什么我们要费尽全力将它们转化成真正的数学论点。一旦它们成为真正的数学观点，它们就不再处于争论之中，至少不会再被了解数学的人们争论了。

其他长小数的展开式

“无穷长小数的展开式”意味着什么？我们现在对此已经有一些线索了。这句话意味着这件事是关于射击挑战的。除非增加的新的小数位上的数字是0，否则每次当你将一个小数位添加到一个数字上时，你就对这个数字做了一点点改变。但是当你添加的小数位越来越多时，你对这个数字做的改变也变得越来越小。在某一时刻，这个变化量会变得非常微不足道。那时你就知道你将会持续击中邪恶对手的射击靶，无论他们把射击靶设置的多小。

对于循环小数，我们可以根据循环节来展开小数。循环节不必须从第一位小数就开始，循环位数多长都可以，但是循环节必须是重复出现的。下面有一些循环小数的例子：

$$0.111111\cdots \qquad 0.\dot{1}$$

$$0.131313\cdots \qquad 0.\dot{1}\dot{3}$$

$$0.18640278278278\cdots \qquad 0.18640\dot{2}7\dot{8}$$

右边一列是更为清晰的书写小数的方式，这样写我们就能完全看清楚小数的哪个部分是重复的。

关于无理数的一个奇怪的难题就是它们根本没有小数位的展开规律。它的小数数字都是随机的，从来不重复，哪怕是循环节非常长的重复也没有。那么我们如何判断这个数字的极限或者说它对应的靶心是什么呢？这是一个好问题。答案是，我们不能做这样的判断。

在这里，我们不得不返回到起始点重新开始。这种情况会时不时地在数学领域发生，并且可能会让你感到反胃。有时当我乘坐轮船尤其是小船时，我会感到非常不舒服。除非我做一些调整，让自己同步地按照船随波浪摆动的方式摆动。一旦实现了同步，我就会觉得好多了。（然而在主题乐园玩那种大型秋千项目时我根本无法完成这种调整。那些项目只会让我感觉到恶心。）但是，一旦我适应了船上的摇摆幅度，回到陆地又会让我不舒服了。我通常把自己调整得太过适应船的摇摆，以至于上岸后变得无法适应固定的地面。

下面我们就返回起始点来适应新的射击靶。之前我们邪恶的对手在圆心周围设置了射击靶，而我们已经通过了各种各样的考验。这一次，我们并不知道靶心在哪里。我们只有一个越变越小的射击靶。只要我们能像以前一样持续击中射击靶，我们就赢了。我们赢得了什么？我们赢得了一些认知，即虽然我们并不知道靶心在哪里，但是我们知道在某处肯定是有一个靶心的。

无理数的无穷长的小数展开式就属于这种不知道靶心在哪儿的情况。我们知道，随着展开式变得越来越长，数字将会逐步稳定下来，并且持续击中一个很小的射击靶。但是我们永远也无法找到靶心在哪儿。例如，我们来考虑一下π。有一个显而易见的事实是，任何圆形其周长和直径的比值都是相同的。这个比值就是一个无理数，我们称它为π。

但是，π是哪个数字呢？我们并不知道。我们已经获得了π的非常长的小数展开式（大概已经有10万亿位了）。但是这并不意味着我们已经知道了π的完整的小数展开式。根据西蒙·普劳夫（Simon Plouffe）在1995年提出的一个算法，我们可以算出π在任何一个小数位上的数字而不需要知道它前一位是什么数字。即便如此，我们仍然不能立即知道π的整个小数展开式。

我们知道，通过写出非常长的小数展开式，我们能够击中任何一个无穷小的射击靶。比如说，当我们在小数点500万位之后再添加小数位数时，整个数字几乎就是不变的——但只是“几乎”，它仍然是有变化的，而我们并不知道它射向的靶心到底是什么。这就是关于无理数的一个非常奇怪的事实。无理数无处不在，而我们却真的无法说出它们是什么。除非我们用越来越小的射击靶来迂回地表达它们。π恰好可以用非小数的方式来准确表述，因为π就是任何一个圆的周长与直径的比值。同样，$\sqrt{2}$也可以表述成平方为2的正数。然而，绝大部分无理数不能用这种方式来表述。那么我们应该怎么说它们是什么呢？这是一个非常难回答的问题。

什么是实数？

怪不得我们花费了那么长时间来确定实数。在研究实数之前，我们能够非常容易地说出之前的几种类型的数字是什么。我们没费太大的力气就研究了自然数、整数和有理数。而一旦你开始考虑无理数，情况就变得非常复杂。

在数学的历史上，有一个令人兴奋的瞬间：两个不同国家的人以不同的方法几乎在同一时间解决了实数的问题。似乎冥冥中有种力量让世界上不同地域的数学家们能在同一时间解决相同的问题。我并不是想说世界上存在上帝或者造物主。也许只是人们对数学的研究发展到了一个特定阶段，时机成熟了，然后一切就自然地发生了。1872年，康托尔和戴德金各自以完全不同的方法回答了什么是实数这个问题。

康托尔使用的方法就是我们前面描述过的缩小射击靶的方法。人们经常会将康托尔的这个成就和柯西联系在一起，因为缩小射击靶的方法是基于后者的理念提出的。而康托尔将这个方法应用于构建实数的过程中。戴德金构建实数的方法更像是在有理数之间寻找空隙。这两种方法最终都得出了相同的结论，即实数作为数字看起来完全不同于我们曾经思考过的任何一种数字类型。这看起来就像我们先前构建无穷所得出的结论。无穷最终被定义为某个集合中数字的数量，虽然我们并不知道这个集合是什么。

按照康托尔的方法，实数可以通过找出满足射击测试条件的有理数数列来构建。下面是一个著名的例子。我们可以用有理数数列$\frac{1}{}$来定义无理数e。我们知道，n！的定义是：

$$n\times(n-1)\times(n-2)\times\ldots\times3\times2\times1$$

在这里，按照数学的惯例，我们令n从0开始，并且定义0！为1。

定义0！为1的一个合理解释如下：n！就是你播放iPod上的n首歌曲能够采用的不同播放顺序的总数。如果你的iPod上只有0首歌曲——就像我的新手机一样，那么你只有一种播放歌曲的方式：什么都不听。

这里我们考虑的数列如下所示：

$$\frac{1}{}、\frac{1}{}、\frac{1}{}、\frac{1}{}、\frac{1}{}、\frac{1}{}\ldots$$

然后我们持续将它们全部相加直至永远，我们就能通过邪恶对手的射击靶考验了。虽然我们并不真的知道靶心在哪里，但是我们知道它就在某处。我们将这个靶心称为e。

一个奇迹般的事实是：我们同样可以通过函数ex得到一个更广为人知的e的定义。函数ex符合下述两个特征：

- 这个函数的每一点的变化率都是它本身。也就是说，如果你画出这个函数的图，那么图上任意一点的斜率就是那一点对应的y值。

- 0点处的y值以及变化率都是1。

那么，数字e就是y值（及变化率）为1的那一点。

这里还有另一个例子。这次我们将下面这些有理数相加：

$$4、-\frac{4}{3}、\frac{4}{5}、-\frac{4}{7}、\frac{4}{9}、-\frac{4}{}\ldots$$

这么做也同样能让我们通过邪恶对手的射击考验，而同时我们还是不知道靶心是什么。这次我们把这个靶心称为π。靶心是π这个事实是一个奇妙

的、可被证明的结果，但我们必须使用一些非常复杂的运算技巧才能证明它。

为了证明这些复杂的抽象对象具有数字的特征，我们需要做非常多的工作。我们需要证明如何将它们相加，如何将它们相乘，并且需要证明在数字中正确的规则对于这些事物也是成立的，比如加法和乘法的顺序并不会影响运算结果、在等式的两边同时减去某数而等式仍然成立，等等。令人诧异的是，将一些我们用直觉就能理解的事物转化成缜密的数学竟然需要如此复杂的过程。这使得数学让一些人觉得沮丧，让另外一些人觉得重要，而让其他人觉得毫无意义。我们本能的直觉是如此强大，以至于我们用自己的大脑来理解它是如此困难，这是一件多么迷人的事。这并不意味着我们不应该尝试，虽然我们也很满意于像康托尔和戴德金这样伟大的数学家已经帮我们解决了这些问题，因为这样我们就能惊叹于他们的方法，而不必非要自己去尝试解决难题。

16　古怪的终极难题

我们对无穷小的新认知带来了一些奇怪的事实。无穷的事物和有限的事物开始以奇怪的方式混杂起来。

首先，我们来介绍一下如何使用有限的饼干面团制作无穷的饼干。你可以先做出第一块饼干，然后只使用前一块饼干一半的面团用量做出第二块饼干（见图16–1）。

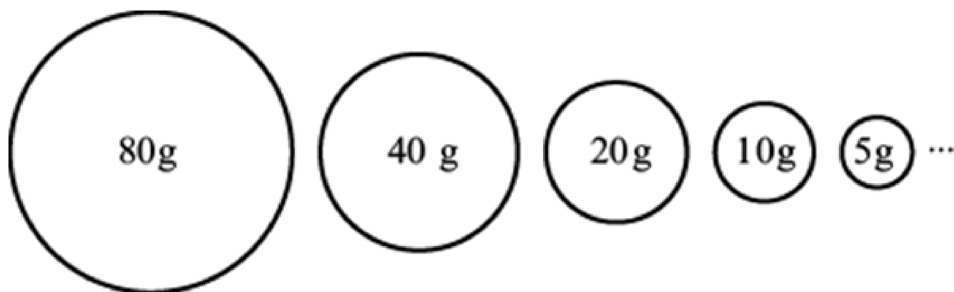


图16–1

下一块饼干使用的面团量又是这次的一半，然后再减半，以此类推。只要第一个饼干没有大到用完一半面团或者更多，你将永远无法用完面团。（不过，最近一段时间中流行的似乎是，用掉所有面团制作一个大饼干然后放在烤盘中烘烤。）你将会拥有无穷数量的饼干，唯一的问题是饼干会变得无穷小。一旦它们小过一个特定尺寸，你就无法再看见它们了。

如果我们让饼干变小的速度稍稍慢一些呢？这次我们不再采用每块饼干的面团用量是前一块饼干一半的这种方式。你可以重新回到第一块饼干，然后使用比它用掉的面团稍稍小一点儿的比例。比如说，第二块饼干用的面团是第一块饼干的$\frac{1}{2}$，下一块饼干是第一块饼干的$\frac{1}{3}$，再下一块是第一块饼干的$\frac{1}{4}$，然后是$\frac{1}{5}$，以此类推（见图16–2）。

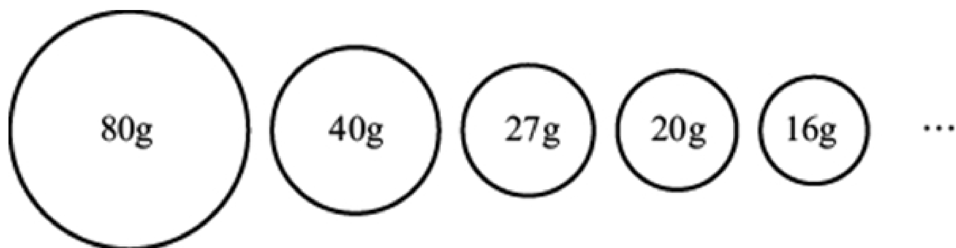


图16–2

问题是，这次你将需要无穷的面团。

另一种情况下，你可以让第二块饼干的半径是第一块饼干半径的$\frac{1}{2}$，下一块饼干的半径是第一块饼干半径的$\frac{1}{3}$，再下一块的半径是第一块饼干半径的$\frac{1}{4}$，然后是$\frac{1}{5}$，以此类推（见图16–3）。

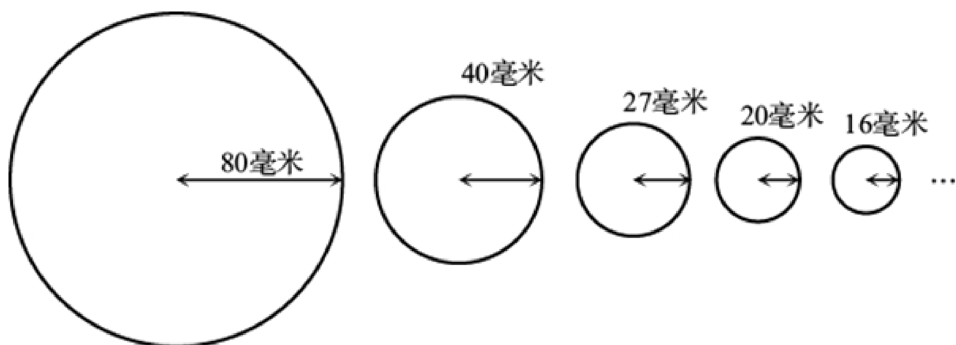


图16–3

这次你可以用有限量的面团来完成。但是如果你将这些饼干排成一条长线，它们将延伸至无穷远。

这又是什么情况？

调和数列

现在我们考虑下面一系列数字：

$\frac{1}{2}$、$\frac{1}{3}$、$\frac{1}{4}$、$\frac{1}{5}$、$\frac{1}{6}$、$\frac{1}{7}$…

我们把这个数列称为调和数列。它与音乐中的和音有关。它是某个音的和音的波长，表达为基音的比值。例如，如果你在小提琴上不按下任何手指，直接拨动G弦，你就会得到G音。通过用手指按下琴弦缩短琴弦真正震动的部分，你就能演奏出更高的音。但是你只能通过在某个特定位置轻轻触碰琴弦来演绎泛音，而不是完全将琴弦压下。泛音听起来比普通的音更加空灵和飘逸。

如果你轻按琴弦上端的$\frac{1}{2}$处，你得到的泛音将比主音G高出八度。如果你轻按琴弦上端的$\frac{1}{3}$处，你将得到比主音G高出八度加五度的D音。轻按琴弦上端的$\frac{1}{4}$处，你将得到下一个泛音；轻按$\frac{1}{5}$处，然后$\frac{1}{6}$处，你都会得到不同的泛音。你无法在小提琴琴弦上得到更多的泛音了，然而你可以在大提琴上弹奏更多的泛音，因为大提琴的琴弦更长，所以我们有更多的空间来寻找泛音。

铜管乐器演奏者可以完全不使用手指而只通过改变嘴形和用力程度来演奏不同的音。黄铜乐器的基音更像是降B调。泛音会比基音高相同的间隔。高八度，接下来再高五度，以此类推。

调和数列在数学和音乐中具有相同的重要性。奇怪的事实是，如果我们持续把这些分数相加到“永远”，其结果永远不会击中任何射击靶。也就是说，自作聪明的人永远都是赢家。这些分数的加和将趋于无穷。我们在第11章讨论增长比较慢的事物时曾经提到过这个事实，只是那个时候我们没有证明它。

如果我们尝试赢得射击挑战，就会发生下面的情况：即使我们暂时击中了射击靶，最终我们还是会再次脱靶。随着我们将更多的分数加在一起，总数会持续变大。无论我们每次添加的量看起来多么微不足道，这个数列的总和都将趋于无穷。请注意一个事实：总和持续增大并不自动意味着我们在射击挑战中失败了。毕竟当我们加$\frac{1}{2}$、再加$\frac{1}{4}$、再加$\frac{1}{8}$，以此类推的时候，这个总和也是持续增长的。只是它增长得足够慢，我们不会让它超过1。然而对于调和数列，虽然它增长得也很慢，但是这一速度也已经足够快了，以至于无论我们试图给它设置什么边界，最终这个总和还是会超越那

个边界。

下面是一些例子。

我们来试试这个数是否会超过1。这个例子和蛋糕的例子不同。在蛋糕的例子中，我们每次吃剩下蛋糕的一半。我们无法吃掉整个蛋糕，直到我们几乎到达“永远”。现在我们试试能否将调和数列中的足够的数相加得到某个大于1的数。我们可以这样做：

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}=\frac{6}{12}+\frac{4}{12}+\frac{3}{12}$$

$$=\frac{13}{12}$$

这就已经大于1了。

我们来看看是否能让数列的加和结果超过2。这一次，我们借助电子数据表来解决这个问题。我的电子数据表告诉我，一旦加到$\frac{1}{}$，我得到的加和结果就将超过2，也就是：

$$\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}+\frac{1}{9}+\frac{1}{10}+\frac{1}{11}>2$$

既然我已经借助电子数据表来计算了，那我们就来看看什么时候数列加和会超过5。这个过程进行得真的很慢，但是最终在加到$\frac{1}{227}$的时候加和结果超过了5。

最初我想要看看什么时候数列加和会超过10。但是这花费的时间实在太长了，所以我放弃了。因此，如果我们想要证明这个数列的加和趋于无穷的话，我们需要一个更好的方法。首先，“用电子数据表来证明”真的不是一个有效的数学手段。而且即便它是一种有效的手段，也存在着其他问题：只是用它来证明数列加和大于10就已经花费太长的时间了，我们怎么能依靠这种方法证明加和趋于无穷呢？

下面是一个更加狡猾而聪明的方法。我们将调和数列中不同数量的分数组合在一起，分数数量分别是1个、2个、4个、8个，以此类推。

- 第一组分数里只有$\frac{1}{2}$，也就是调和数列中的第一个数。

- 第二组里是接下来的两个分数，$\frac{1}{3}$和$\frac{1}{4}$。

- 第三组里的分数数量是前一组的两倍，也就是再接下来的4个分数：

$$\frac{1}{5}、\frac{1}{6}、\frac{1}{7}、\frac{1}{8}$$

- 第四组里的分数数量是前一组的两倍：

$$\frac{1}{9}、\frac{1}{10}、\frac{1}{11}、\frac{1}{12}、\frac{1}{13}、\frac{1}{14}、\frac{1}{15}、\frac{1}{16}$$

所以在做加和的时候，我们有如下的组合：

$$\underbrace{\frac{1}{2}}_{\text{第一组}} + \underbrace{\frac{1}{3}+\frac{1}{4}}_{\text{第二组}} + \underbrace{\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}}_{\text{第三组}} + \underbrace{\frac{1}{9}+\frac{1}{10}+\frac{1}{11}}_{\text{第四组}}$$

现在我们暂停一会，全面思考一下。我将要证明每一个组合的加和结果都会大于$\frac{1}{2}$。除了第一组，因为第一组只有$\frac{1}{2}$，所以加和结果就是$\frac{1}{2}$。

我们很懒（或者更确切地说，我们喜欢保存脑力），所以我并不打算真的将这些组合一一做加和。我只是要证明每个组合里的分数相加得到的数都会大于$\frac{1}{2}$。以第二组为例，其包含$\frac{1}{3}$和$\frac{1}{4}$。这两个数的每一个数都大于或者等于$\frac{1}{4}$，因此总和将会大于或者等于$2\times\frac{1}{4}$。因为这个组里只有2个分数，每一

个都大于或者等于$\frac{1}{4}$，而$2\times\frac{1}{4}=\frac{1}{2}$。

$$\underbrace{\frac{1}{3}+\frac{1}{4}}_{>\frac{1}{4}+\frac{1}{4}=\frac{1}{2}}$$

我们并没有将分数加在一起，只是证明了$\frac{1}{3}+\frac{1}{4}>\frac{1}{2}$。这是一个更加高级的方法。关键是我们在这里提出了一种归纳的方法，从而帮助我们解决了任意组分数的加和问题。而把分数转换成同分母分数相加或者使用电子数据表，都不能达到这样的效果。当你将归纳的方法用于基础的例子时，这些方法往往看起来很蠢。这就是学习数学很难的原因之一，因为基础的例子常常使数学看起来愚蠢、没有意义，或者装腔作势。

我们接着往下看第三组。这里包含4个分数。这4个分数中最小的是最后一个，$\frac{1}{8}$。因此所有其他的分数都会比$\frac{1}{8}$大，这意味着它们的加和结果将会大于$4\times\frac{1}{8}=\frac{1}{2}$。

$$\underbrace{\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}}_{>\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{4}{8}=\frac{1}{2}}$$

现在，我们来看一下第四组。这个组合包含8个对象，其中最小的一个是$\frac{1}{16}$。因此这组分数的总和将会大于$8\times\frac{1}{16}=\frac{1}{2}$。

$$\underbrace{\frac{1}{9}+\frac{1}{10}+\frac{1}{11}+\frac{1}{12}+\frac{1}{13}+\frac{1}{14}+\frac{1}{15}+\frac{1}{16}}_{>\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{8}{16}=\frac{1}{2}}$$

我们可以永远这样继续下去。下一次我们会选择16个分数作为第五组，最后一个分数是$\frac{1}{32}$。每一个分数都会大于或者等于$\frac{1}{32}$，因此总和就大于$\frac{16}{32}=\frac{1}{2}$。

接下来，我们选择32个分数作为第六组，最后一个分数是$\frac{1}{64}$。然后我们选择64个分数作为第七组，最后一个分数是$\frac{1}{128}$。以此类推。

比起“持续这样到永远”或者“以此类推”，一个更好的表述方法是，当n为任意数时，第n个组合会出现怎样的情况。在这里，第n个组合有2^{n-1}个分数，最后一个分数是$\frac{1}{2^n}$。该组里的每一个分数都会大于或者等于$\frac{1}{2^n}$，因此其总和大于$\frac{}{2^n}=\frac{1}{2}$。可见答案和我们预期的一样。

因此我们可以看出，如果我们把所有的组合相加，永远加下去，其总和将会大于我们持续将$\frac{1}{2}$与它自身相加直到永远。而如果我们将无穷的$\frac{1}{2}$相加，总和肯定会不停增大。

我们还有另一个方法可以用来考虑这个问题：如果邪恶的对手给了我们一

个射击靶，我们的确能够击中它，但是我们注定会再次脱靶。因为我们最终会添加足够多的数字组合，而每一个组合都将大于$\frac{1}{2}$，所以我们的总和将会变得太大而最终从另一边脱离射击靶（见图16–4）。

这时，邪恶的对手就获胜了。

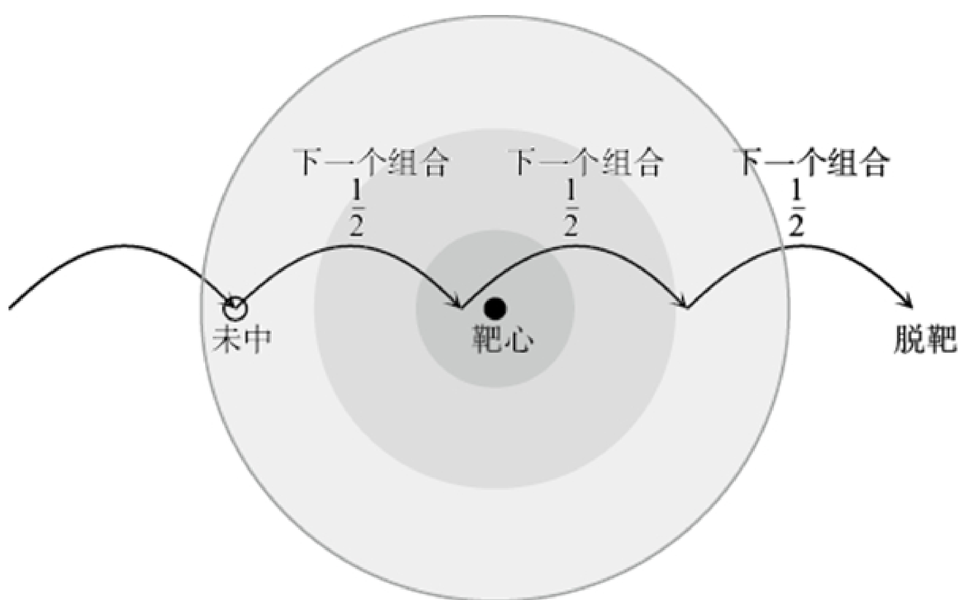


图16–4

柱状图

我们可以用一种柱状图来画出调和数列，如图16–5所示。

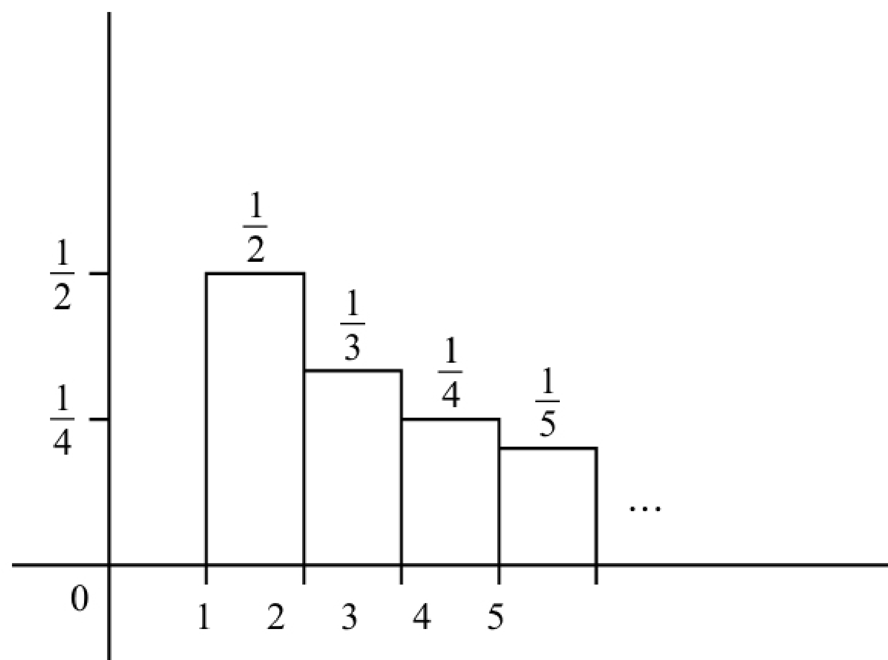


图16–5

图中每个条柱宽为1，相应的面积就是条柱的宽度乘以高度。换言之，对于每个n，它对应的条柱面积为$\frac{1}{n}$。所以所有这些条柱的面积的总和与我们刚刚努力做的调和数列加和的结果是相同的。这意味着图16–5所示的这些条柱最终组成的面积是“无穷的”。其真正含义是，无论我们想到一个多大的数字，我们总会找到一个点，到这点为止所有条柱面积的加和肯定比我们想到的那个数字要大。只要我们沿着这张图向右走到足够远找到越来越大的n，我们总能找到这一点。用我们在第11章讲过的概念，这表明随着n越来越大，条柱的面积将会趋于无穷。

现在我们可以再讨论一下一条曲线下的面积这个问题。如果我们画出$\frac{1}{x}$的图像，那它看起来应该和图16–6一样。

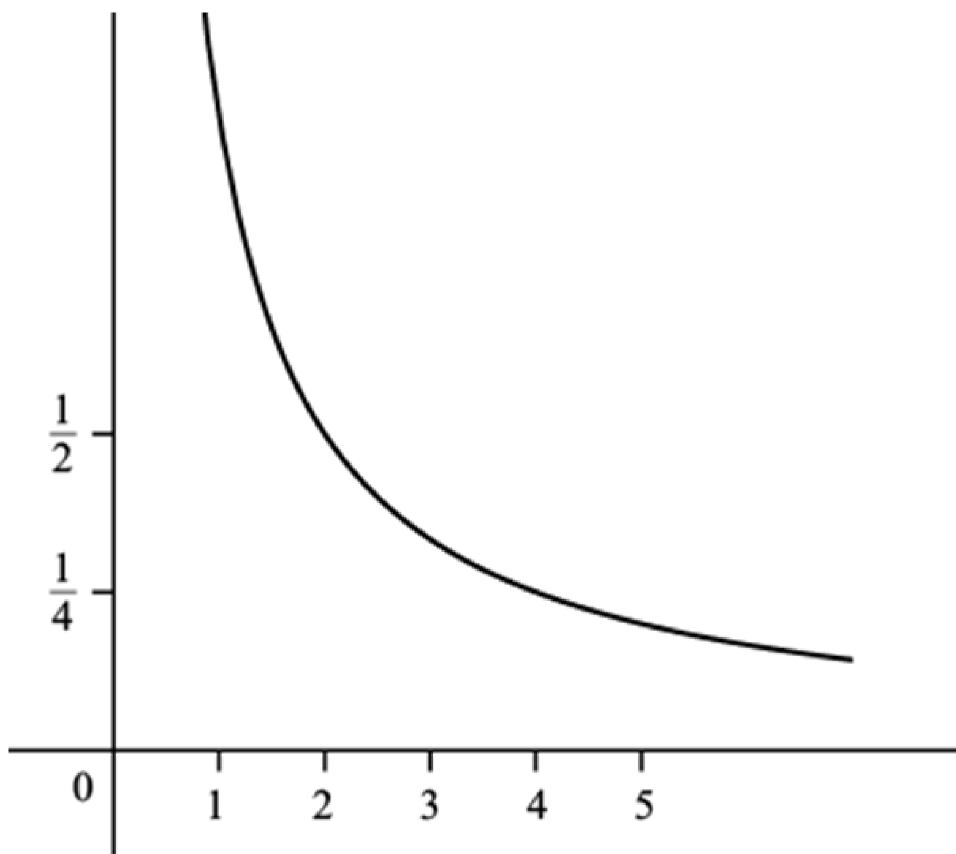


图16–6

首先需要注意的是，当x趋近于0时，$\frac{1}{x}$会变得非常大。我们无法说出当x确切等于0时会发生什么。但是我们可以说当x趋于0时，$\frac{1}{x}$趋于无穷。按照前面的表述方式就是，无论我们想到一个多大的数字，我们总能找到一个足够小的x使得$\frac{1}{x}$比这个数字更大。

另一方面，随着图形向右发展，x会变得无穷大，而$\frac{1}{x}$会变得无穷小。所以

你可能会想，如果我们忽略最开始的部分，也就是$\frac{1}{x}$趋于无穷的部分，那么曲线下的面积应该是有限的。但事实并非如此。

将曲线重叠在之前的柱状图上，我们就可以看到，柱状图紧密地贴合在曲线下面（见图16–7）。

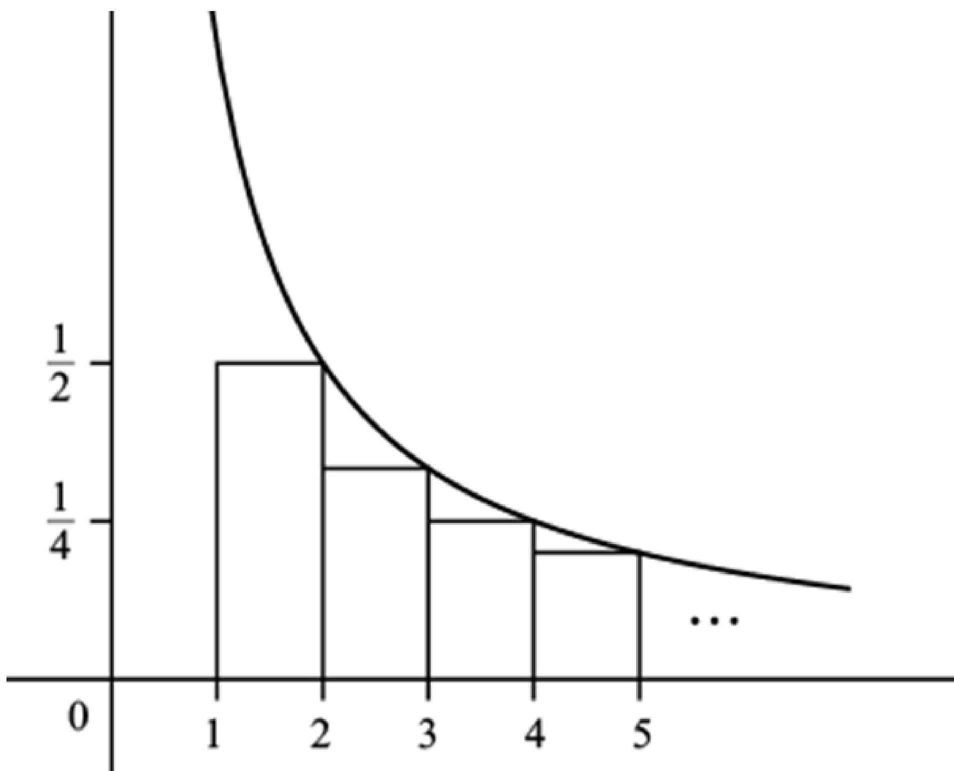


图16–7

我们知道，条柱的面积总和是无穷的。如果你观察条柱与曲线之间的空隙，就会发现条柱并没有填满整个曲线下的区域。换言之，曲线下的面积将比条柱的面积总和更大。如果条柱的面积已经是无穷的事物了，那么曲线下的面积肯定也是无穷的。

事实上，曲线下的面积可以通过自然对数来测量，自然对数记作ln（也就是以e为底数的对数，但是我们现在先不讨论这个话题）。事实上，这正是定义自然对数函数的方式之一。ln b就是这条曲线下从1到b之间的面积。对图16–7的讨论表明，自然对数会缓慢增长至无穷。这和我们在第11

章所做的断言是相同的。

曲线下的面积

在一条曲线下紧紧贴合一个柱状图是求得曲线下面积的关键。这就是所谓的“积分”。这有点儿像在上一章中，我们为了估算一个圆的面积而把它切割成整齐的多边形和三角形一样。现在，我们可以把曲线图切割成矩形条柱，并通过将矩形条柱的面积相加来估算曲线下的面积。矩形越窄，对面积的估算将会越精确。

计算曲线下的面积是推动我们理解无穷小的小事物这个概念的另一个巨大动力。起初，人们尝试使用“无穷小的小”这个长度概念，并且试图解释宽为“无穷小的小”的矩形是什么意思。数学家们很不放心这种方法，而且事实证明他们的不放心是正确的。正如我们之前所说的，从数学的角度讲，相对于我们可以把无穷大与实数相类比，无穷小的小长度比这要难理解得多。

于是，出现了另一个伟大的时刻，又有两个人在相似的时间身处不同的国家用不同的方法解决了曲线下面积的问题。这次是勒贝格和黎曼，但是他们两个人解决问题的方法并不像康托尔和戴德金在构建实数时使用的方法那么接近。黎曼利用垂直条柱解决了这个问题，就像我们上面$\frac{1}{x}$的例子一样。而勒贝格使用的是水平条柱。（这是一个过度简化的说法，但是能够帮助我们大体上明白他们所使用的方法的差异。）勒贝格的方法可能不太直观，但是更加有效。我并不打算在这里讨论他的方法，因为这个方法需要更多的专业技巧。

黎曼的方法是使用垂直条柱提供面积的两个估算值。一个是高估值，另一个是低估值。我们再次以$\frac{1}{x}$的图像为例。用垂直条柱将它切割可得下面的图形（见图16–8）。

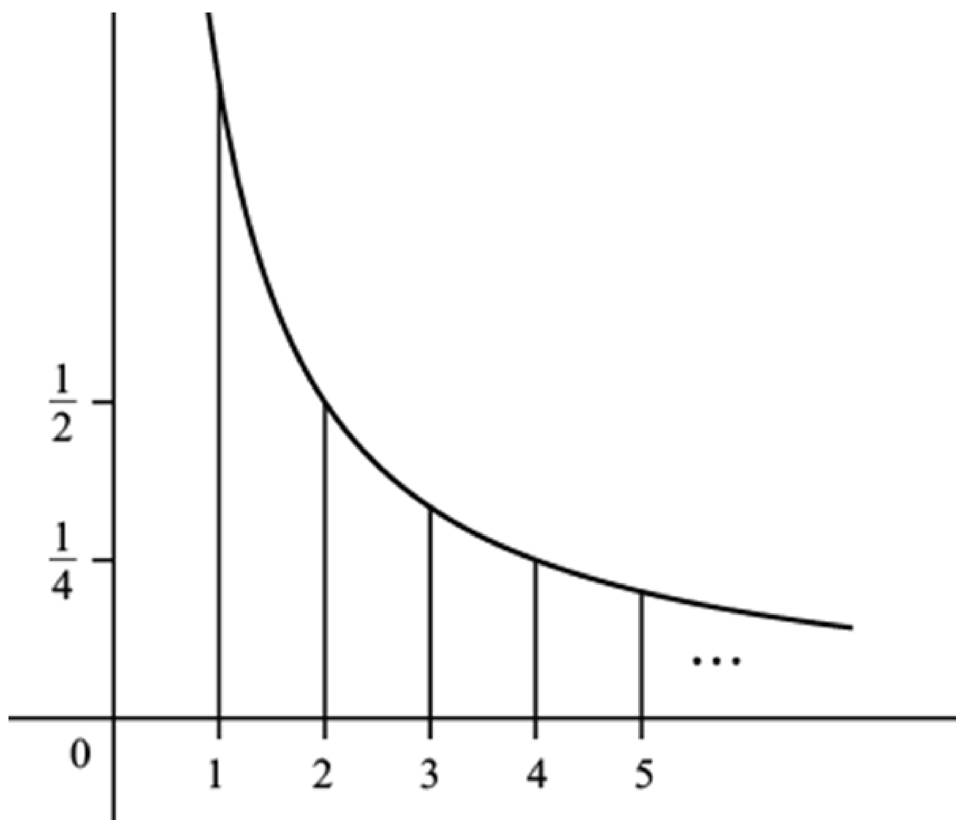


图16–8

这个时候，我们有两种方式来构造矩形。我们可以令矩形的水平上边高于曲线，如图16–9所示。

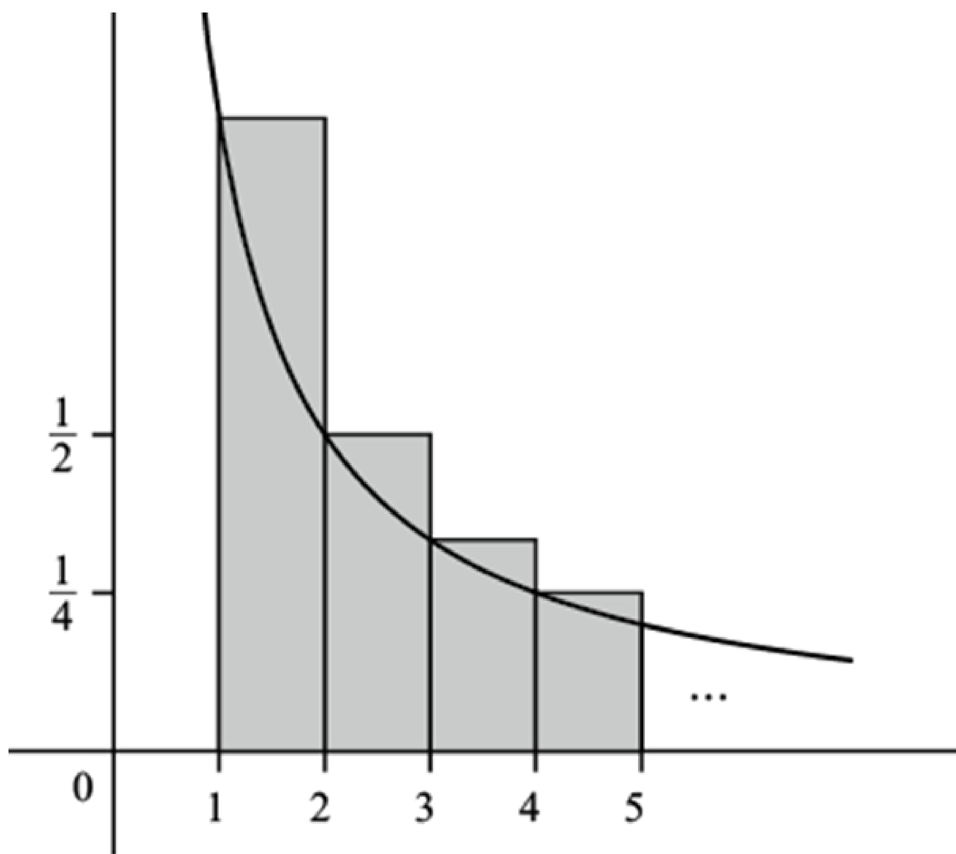


图16–9

或者让矩形的水平上边低于曲线，如图16–10所示。

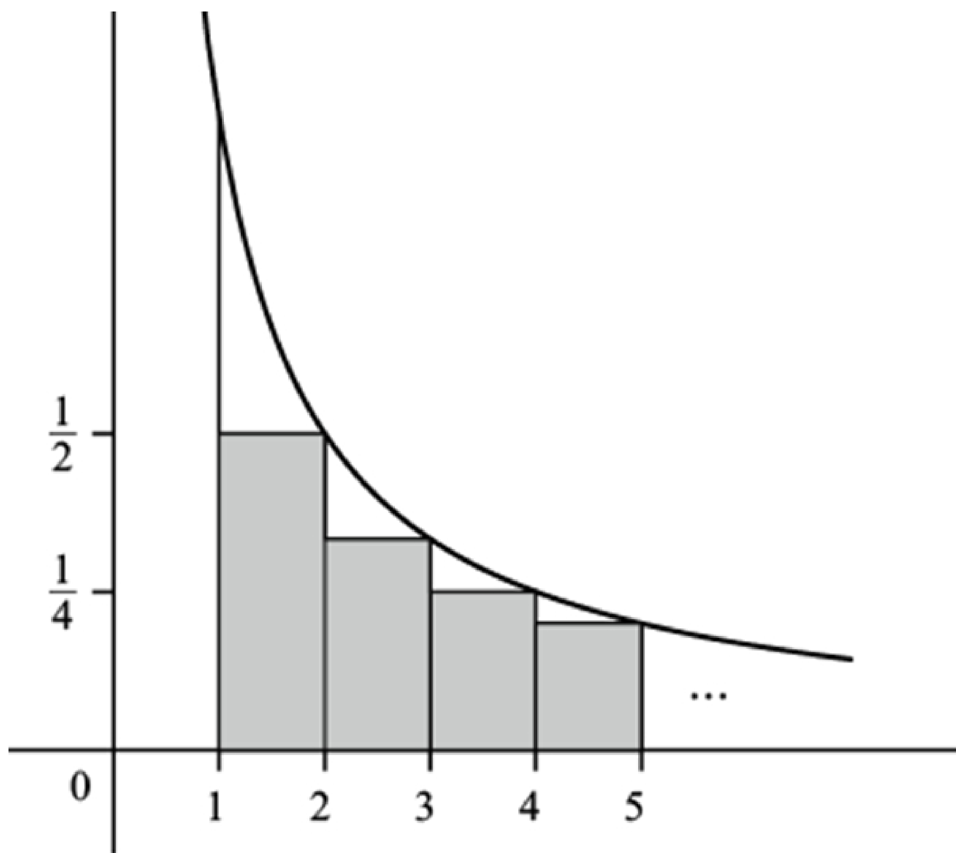


图16–10

这有点儿像我们近似计算圆形面积时用的方法。当时，我们用一个多边形拟合在圆形的外侧，再用一个多边形拟合在圆形的内侧，把圆形夹在这两个多边形之间。现在，如果我们令矩形高于曲线，那么我们所获得面积数值将比曲线下的面积数值更大，所以它是对曲线下面积的高估值。但是如果令矩形低于曲线，那么我们就会漏掉一些曲线下的面积，所以它是对曲线下面积的低估值。如果我们把条柱切割的更窄，这两个估算值都会更接近曲线下面积从而让估算得到优化。例如，如果我们把条柱的宽从1减小到0.5，我们将得到图16–11，面积的高估值将得到“优化”。优化的程度如图16–11中阴影部分所示。

如果我们对低估值进行相同的操作，估算也将得到优化。优化的程度如图16–12阴影部分所示。

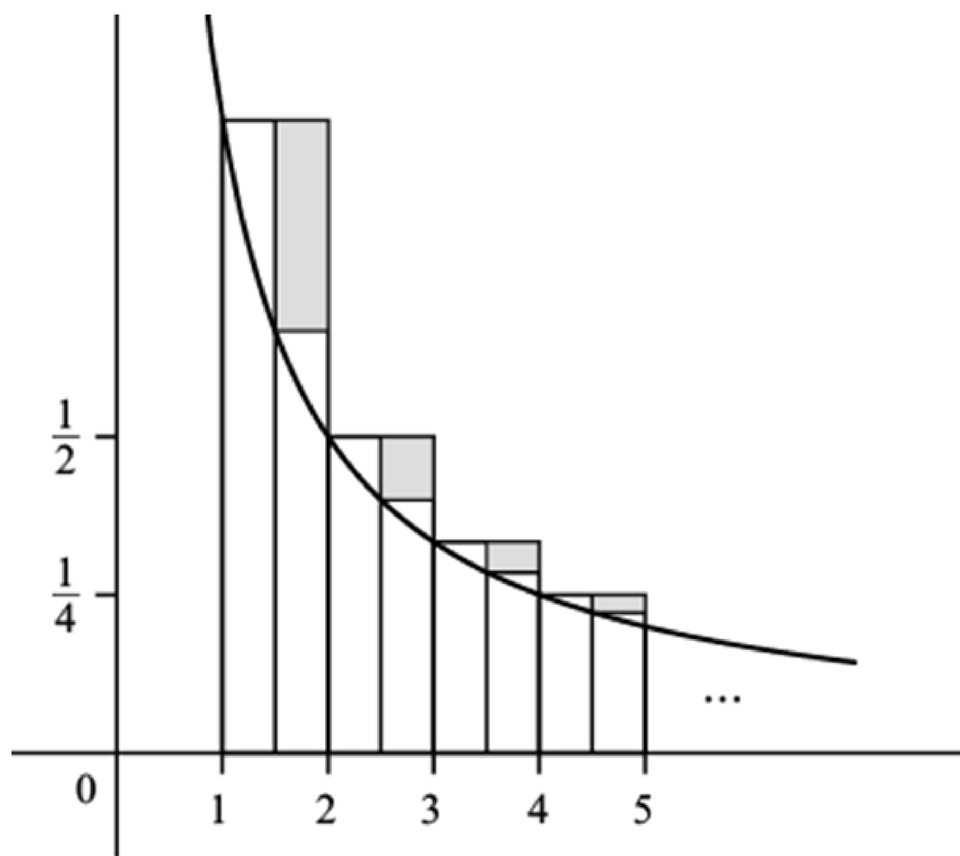


图16–11

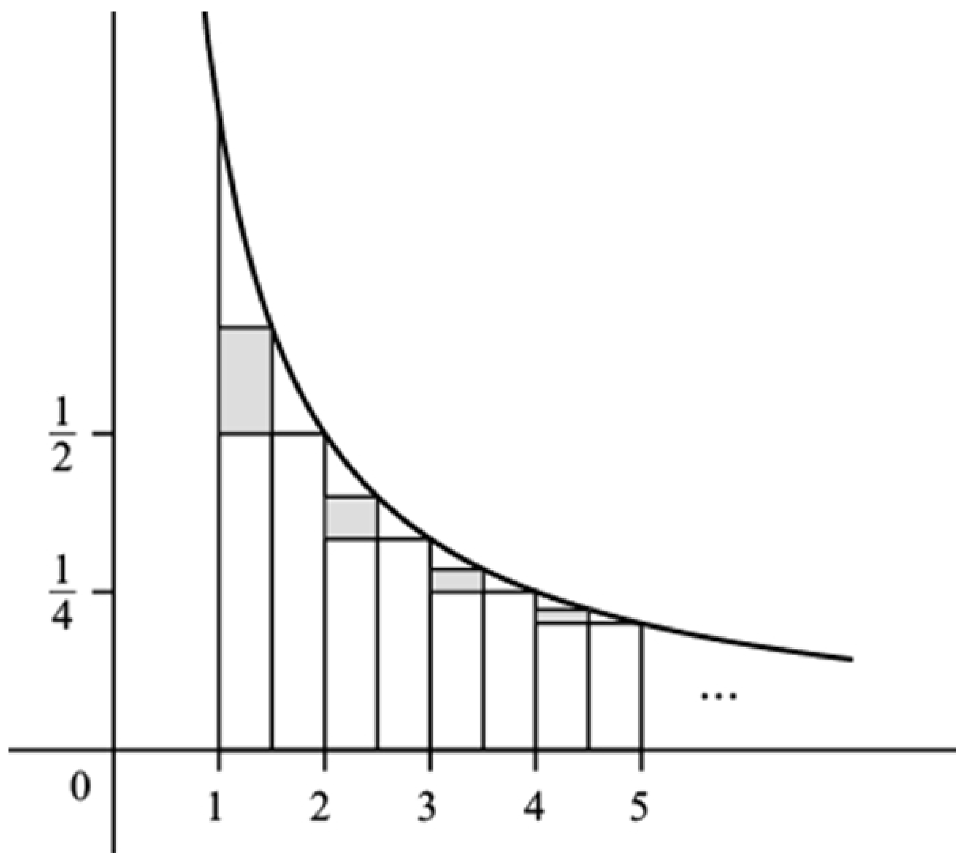


图16–12

真正的曲线下面积就夹在这两个估算值之间。现在的问题是，高估值和低估值是否会“相遇”。它们可能不会真正相遇，但是它们会变得非常接近以至于它们之间的差距已经变得微不足道了。有多微不足道呢？微不足道到足以再次战胜那些自作聪明的人。这就像你和你的朋友一起进行射击挑战，你们必须永远同时击中同一个射击靶，无论这个射击靶有多小。如果你们能同时做到，那么你们就胜利了。这次也一样，你们不需要知道靶心在哪里，虽然它肯定在某处。这次的靶心就是曲线下面积的答案。

当你在学校学习积分时，你通常会把它作为一系列漂亮的公式来学习。如果要求你计算一个特定函数的积分的话，你就会对这个函数依照公式进行一些操作从而得出答案。但是这些公式能够运作的原理就是我们刚刚进行的将图形切割成无线小的小条柱的过程。就像无穷饼干一样，射击挑战帮助我们从数学的角度搞清楚这是什么意思。

终极饼干难题

现在我们回到本章开篇提到的那个最奇怪的饼干难题。在这个难题里，我们用有限的面团制作了无穷的饼干，而且当我们把饼干排在一条线上时它们会延伸到无穷。这怎么可能？

我们制作了一系列的饼干，其中第二块饼干的半径是第一块的$\frac{1}{2}$，第三块饼干是第一块的$\frac{1}{3}$，然后是$\frac{1}{4}$，以此类推（见图16–13）。

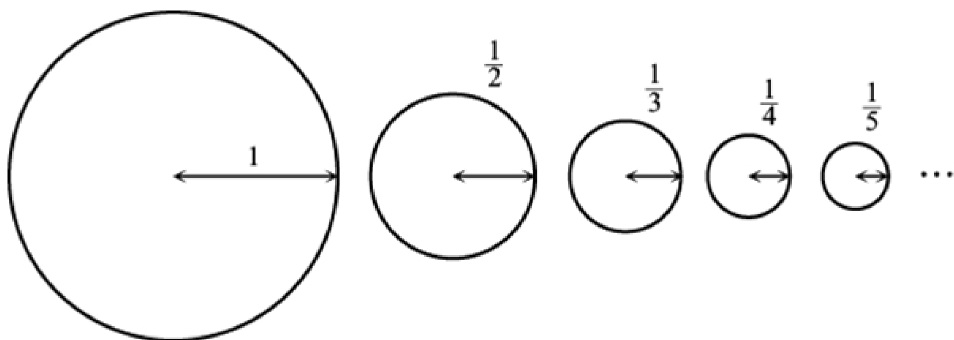


图16–13

当我们把这些饼干排在一条线上时，这条线的长度将是调和数列的两倍。也就是说，其长度是一个已知无穷的两倍，所以它必然是无穷的。但是，为什么饼干所用面团的体积依然是有限的？

我们假设饼干是圆形并且完全平整的。顺便一提，我曾经因为用烤饼进行假设而被听众批评，被指责使用的是工厂生产的烤饼。在此声明，我确实自己制作烤饼，而且我完全清楚在地球上没有任何事物是完全圆而平整的。但是对于数学讨论来说，这是不错的近似！毕竟这并不是什么生死攸关的问题。

总之，每块饼干所用面团的体积等于圆形的面积乘以饼干的厚度。饼干面积的计算公式是πr^2，而第n个饼干的半径是$\frac{1}{n}$。我们记饼干厚度为某个固定量t。随着饼干变得无穷小，将饼干的厚度想象成保持不变的恒量有一些古怪。但是这种近似方法仍然会给我们一个有限量的体积。而如果饼干厚度像半径一样变得越来越小，我们一定也会得到一个有限的总体积。

这种提出一个稍稍有些不切实际的假设的方法是学习纯数学的关键。我们尝试回答一个特定的问题，然后我们可以在尝试过程中做各种各样的假设，只要我们认为这些假设不会对问题的答案产生严重的影响，即使假设让计算不那么“生活化”也没什么关系。如果我们试图回答的问题是，“准确地讲，我们使用了多少面团？”那么饼干全都具有相同厚度这个假设肯定会影响答案。然而，我们只是试图回答一个更简单的问题：“面团的量是有限的还是无穷的？”

有了这些假设，第n个饼干的体积就是πr2乘以厚度t，也就是：

$$\pi \times \left(\frac{1}{n}\right)2 \times t = \frac{\pi t}{n^2}$$

于是我们有了下面这个关于体积的数列：

$$\frac{\pi t}{2^2}、\frac{\pi t}{3^2}、\frac{\pi t}{4^2}、\frac{\pi t}{5^2}、\frac{\pi t}{6^2}\ldots$$

πt的部分是不会变化的。那么我们可暂时只考虑变化的部分，也就是数字：

$$\frac{1}{2^2}、\frac{1}{3^2}、\frac{1}{4^2}、\frac{1}{5^2}、\frac{1}{6^2}\ldots$$

这个数列看起来很像调和数列：

$$\frac{1}{2}、\frac{1}{3}、\frac{1}{4}、\frac{1}{5}、\frac{1}{6}\ldots$$

不同之处在于每一个分数都取了平方。如果做的是方形饼干而不是圆形饼干，我们也会得到这个结果。一个小于1的分数在取平方后会变得更小（因为你等于是取了这个分数的一部分）。所以这些数比调和数列中的分数减小得更快。事实证明，这是判断一个总和会增大到无穷还是稳定于有限的关键：每一个单独的对象都变得越来越小，但是不同的数列变小的速度不一样。调和数列中的数字变小得不够快，所以它的总和是无穷的，而

这次的数列总和是有限的。事实上，这次射击靶的靶心位于$\frac{\pi^2}{6}$–1处。

这很难证明，但是你可以尝试通过在1×1的正方形里排列小正方形来说服

自己。如果我们使用一个边长为$\frac{1}{n}$的小正方形，那么小正方形的面积将是$\frac{1}{n^2}$。我们使用如下边长的正方形来填充1×1的正方形：

$$\frac{1}{2}、\frac{1}{3}、\frac{1}{4}\ldots$$

然后求出下面数列的总和：

$$\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\ldots$$

之后比较它和大正方形的面积数值。大正方形面积为1。

第一步，我们可以把$\frac{1}{2}\times\frac{1}{2}$这个正方形放在$1\times1$这个大正方形的角上，如图16–14所示。

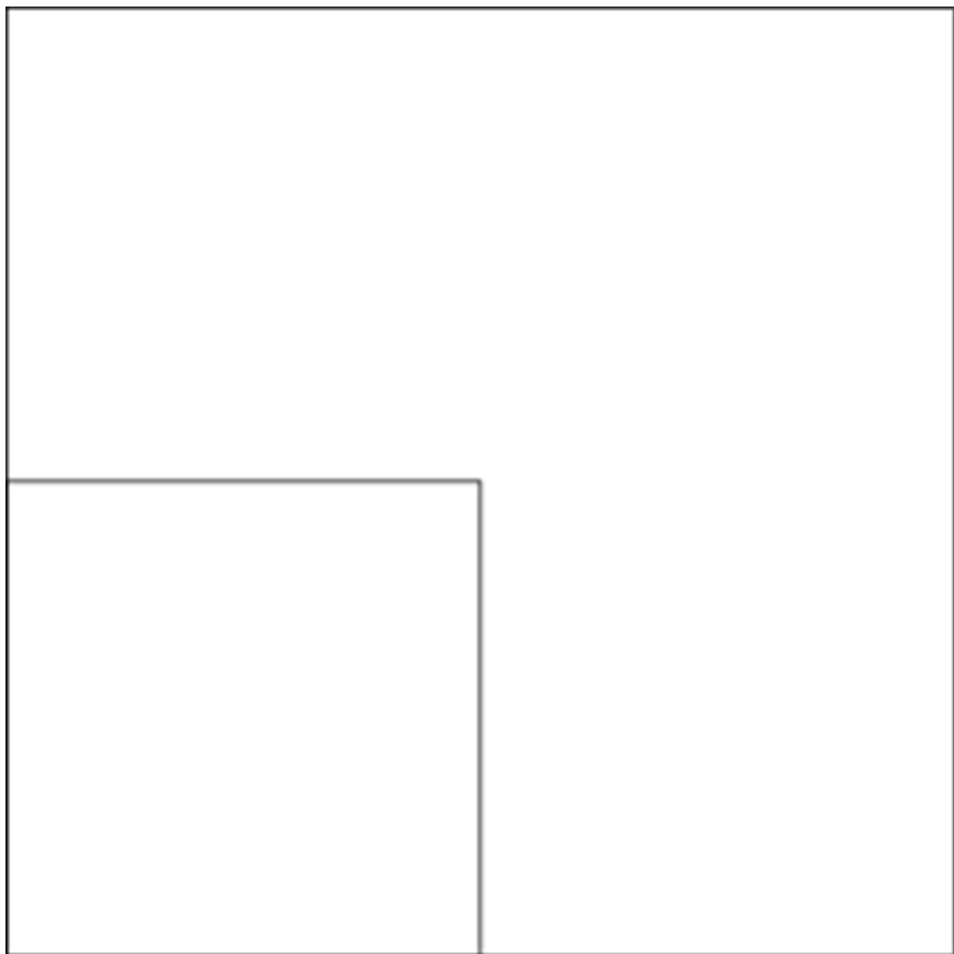


图16–14

接下来，我们放进去一个$\frac{1}{3}\times\frac{1}{3}$的正方形，如图16–15所示。

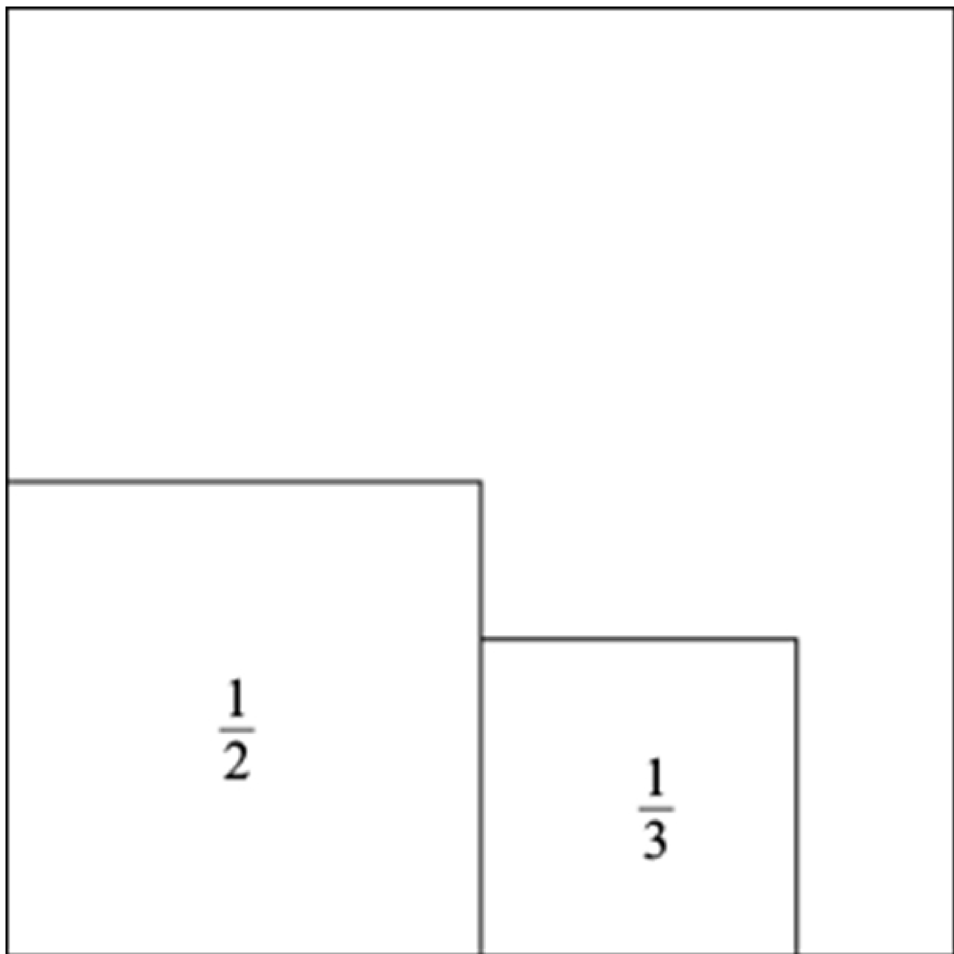


图16–15

现在，我们无法把$\frac{1}{4}\times\frac{1}{4}$的正方形放在右下角了，因为剩下的空间不够了。

剩下的空间的宽度只有$\frac{1}{6}$。但是我们可以像图16–16所示，把它放在$\frac{1}{2}\times\frac{1}{2}$的正方形上面。

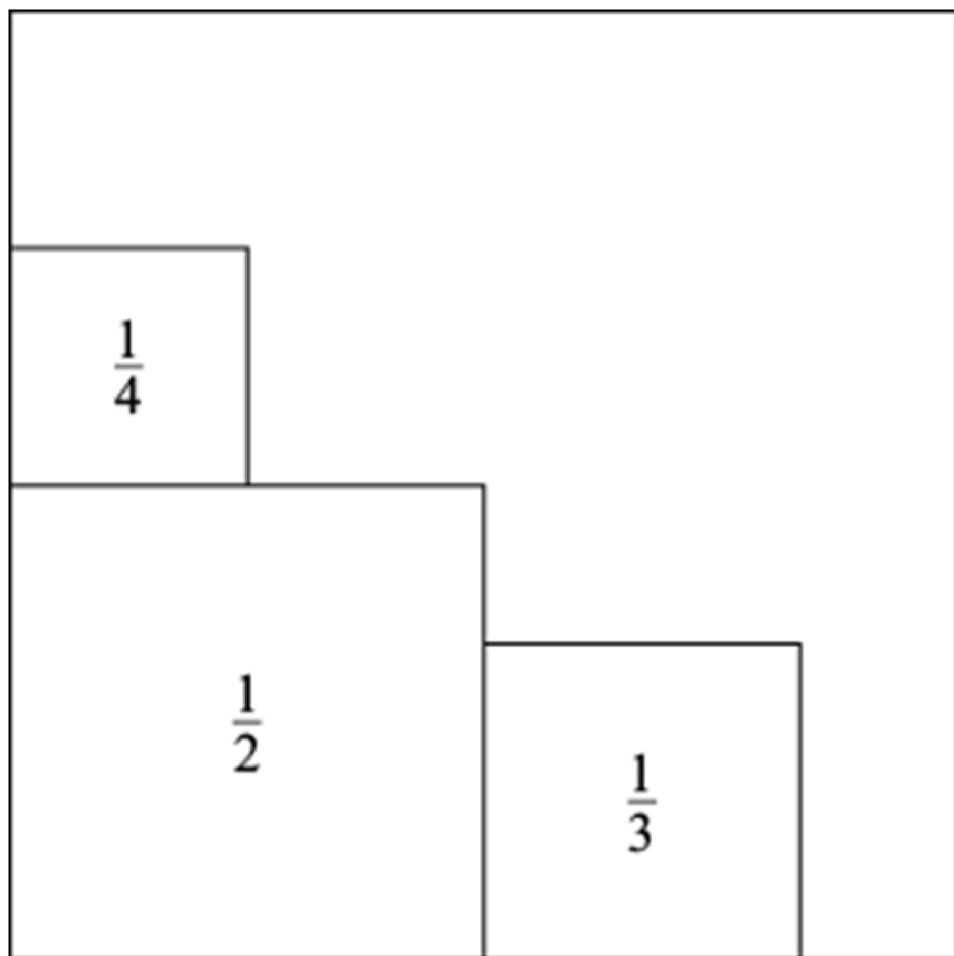


图16–16

我们可以像图16–17所示这样多做几步。

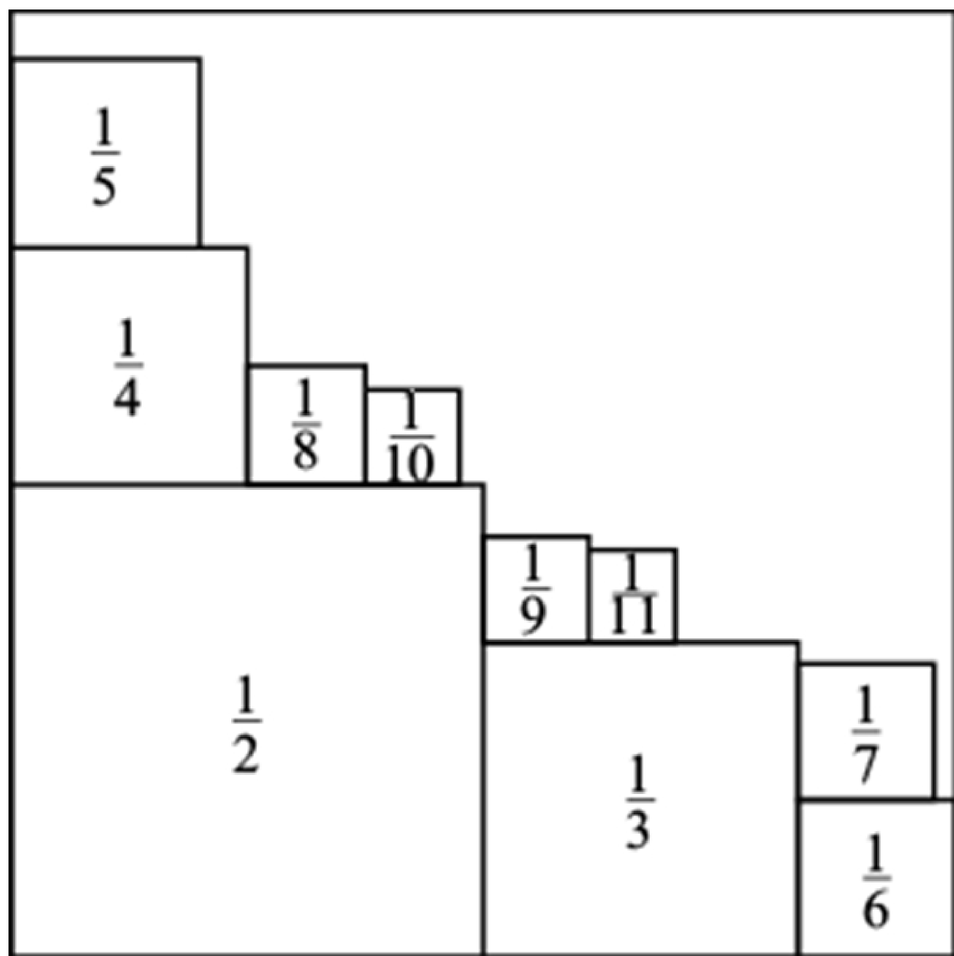


图16–17

如果你亲自试一试的话，你会发现总是会剩下很大的空间让你来放置下一个正方形。这不是一个证明，但它确实让你感觉你好像永远都无法用尽剩余的空间。事实上，你也许也有同感，那就是你根本用不到右上角那$\frac{1}{4}$的正方形（见图16–18阴影部分）。

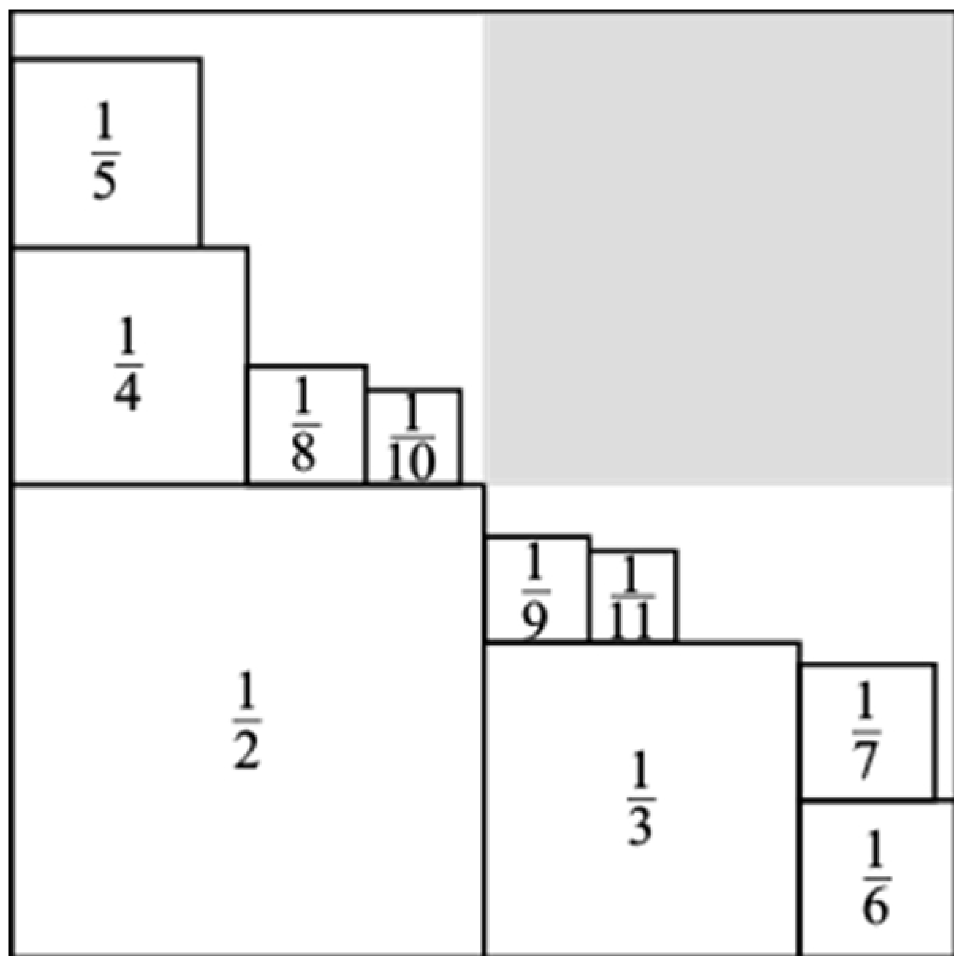


图16–18

这意味着我们可以将所有的小正方形放进$\frac{3}{4}$的面积里。这是正确的。如果你在计算机里输入$\frac{\pi^2}{6}-1$，你会发现它大约等于0.645，小于$\frac{3}{4}$。

如果你还能记起如何进行积分计算，那么通过研究$\frac{1}{x^2}$曲线（见图16–19），你将不难证明总和是有限的。

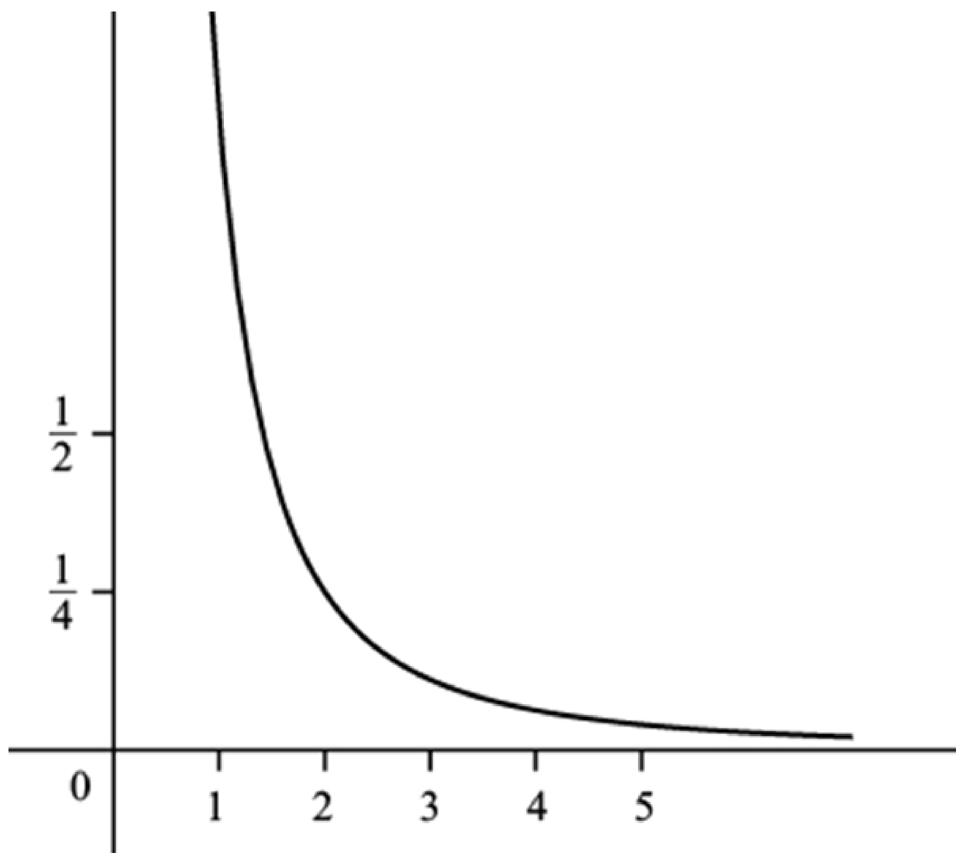


图16–19

如果我们以图16–20所示的方式将曲线切割成垂直条柱，可以看到曲线下的面积将大于我们想要求的总和，也即垂直条柱面积的总和。

如果这条曲线下的面积是有限的，那么我们所要计算的总和肯定也是有限的。如果你还记得如何计算$\frac{1}{x^2}$积分，你将不难算出曲线下从1到b的面积为$1-\frac{1}{b}$，无论b变得有多大，这个面积都会小于1。所以它肯定是有限的。

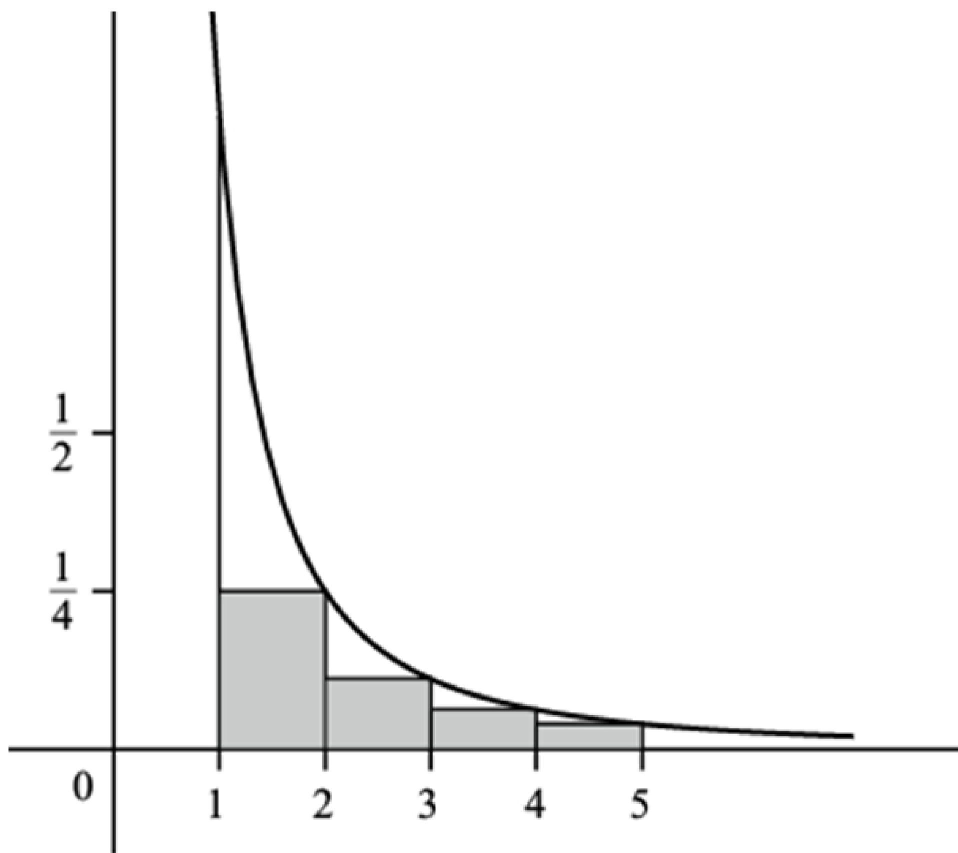


图16–20

我们的饼干的总体积并不等于这个平方版本的调和数列之和，但是我们需要做的就是把这个总和乘以π再乘以饼干的厚度t。因为之前的总和已经是有限的了，所以做完乘法之后得到的数字仍然是有限的。（除非我们拥有的是无穷厚的饼干。很遗憾，我们没有。吃无穷厚的饼干也许是一件美好的事，但是我会变得无穷胖。）另一个考虑这最后一步的方法是，我们可以把每一块圆饼干放进图16–17中相应尺寸的正方形里，而每一个正方形都会留下空隙未被饼干填满。

因此，饼干面团的总体积是有限的。虽然当我们把所有饼干边挨着边地摆成一条线时，它们会延伸无穷远的距离。

难以置信的体积

关于饼干的这个奇怪情形关系到一个更奇怪且更具数学性的例子。存在这样一个事实，那就是如果你选择一个有无穷面积的面并且让它在空中旋转，那么你还是只扫过了一个有限的体积。在空中旋转一个形状这个理念有点儿像使用一个陶工轮盘。轮盘是旋转的，所以无论你用手做出什么形状，最终这个形状都会绕着陶罐围上一圈。我常常对此感到着迷，但是从来没有机会尝试。这让我想起了用纸剪雪花的神奇过程。将纸折叠起来之后，只剪一下就相当于对称地剪了8次或者16次。陶工轮盘就是这个过程的一种平滑版本。

想象一下，你有一个那种制作大型泡泡的塑料环，通过在空中拖拽它你就能制作一个大的泡泡。如果你拖拽着它转一整圈，你将会获得一个环形甜甜圈的形状。这个形状的正式称呼是圆环体。你可以想象拿着其他形状的泡泡制作环，然后在空中转一大圈。你会做出一个形状，它从一个横截面看是圆形的，从另一个横截面看是环形的。

我们可以用下面这个形状来做这件事（见图16–21）。

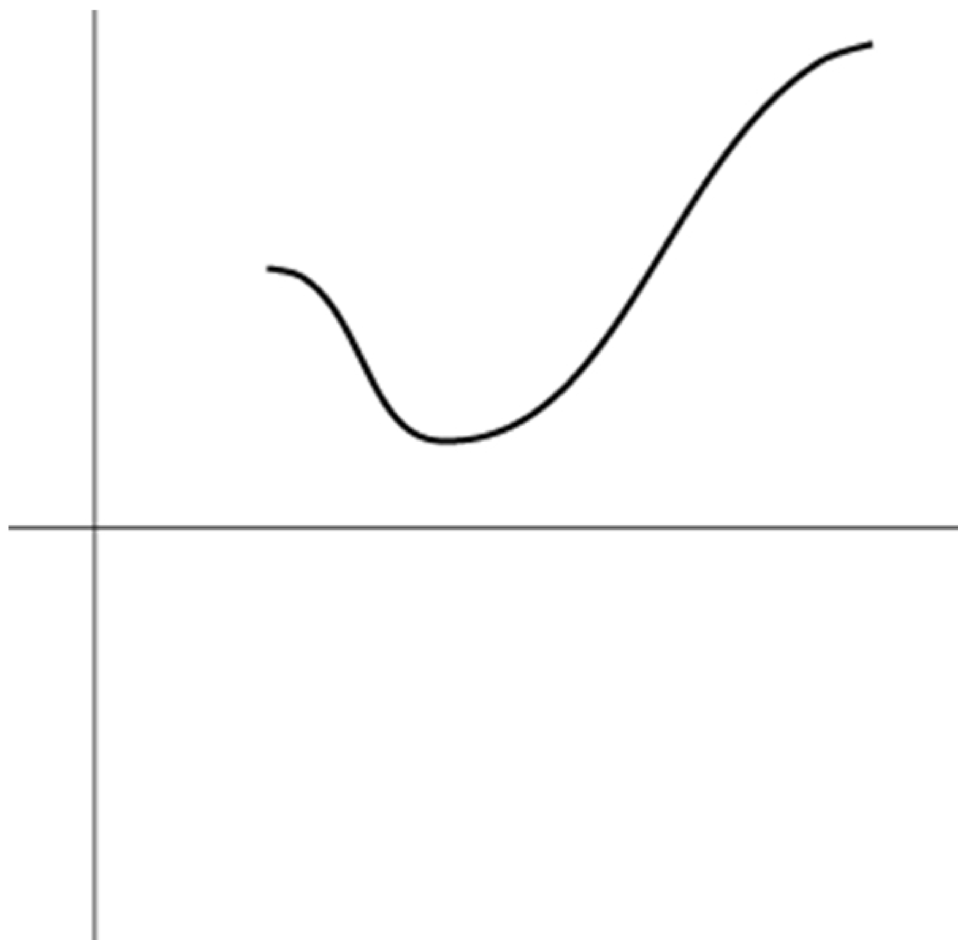


图16–21

沿着x轴一直旋转它，你就能制作出一个花瓶状的立体图形（见图16–22）。

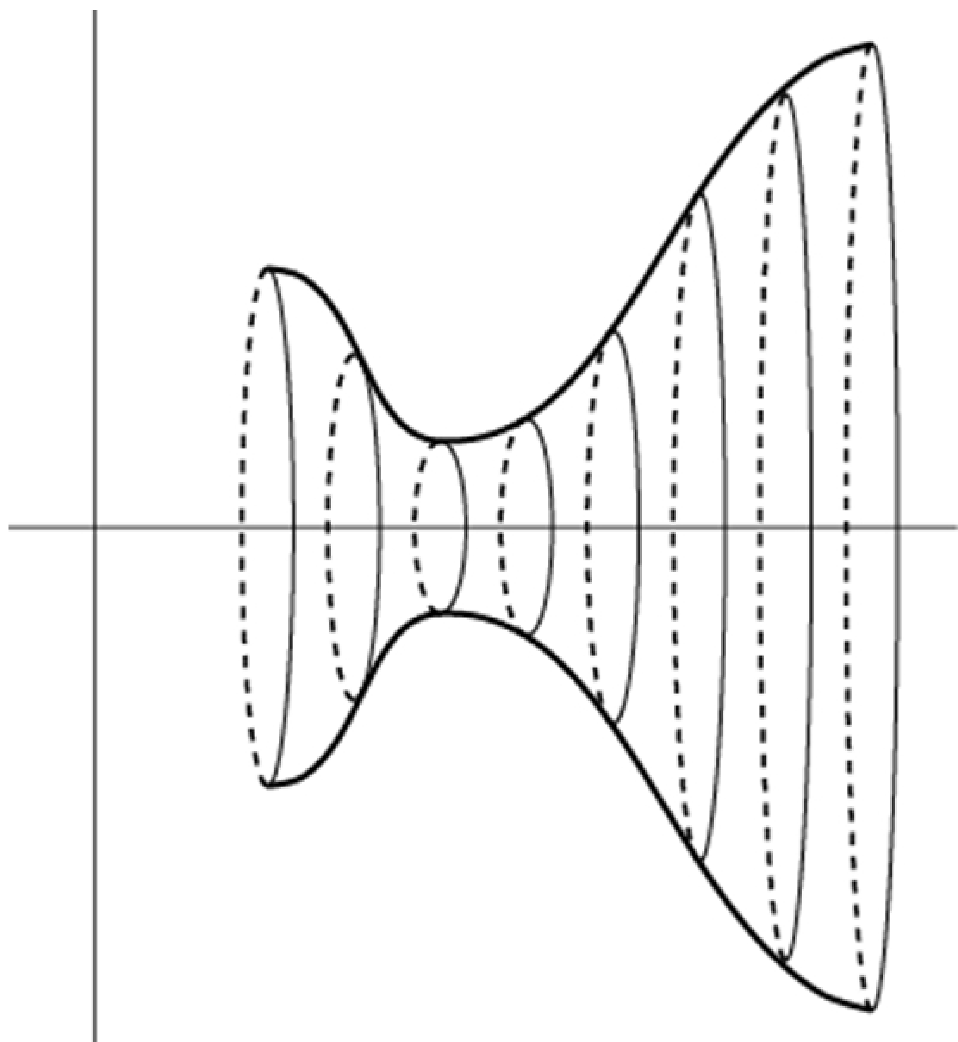


图16–22

我们把这类形状称为“旋转体”，因为它们都是通过绕着某个轴做一次完整的旋转制作出来的。

现在，如果我们选择$\frac{1}{x}$的曲线，沿着x轴旋转它，我们会得到下面的形状（见图16–23）。

这个形状看起来有点儿像把我们在前面的例子中制作的饼干们竖着堆叠起

来。任意n处对应的图形半径为$\frac{1}{n}$，因此我们在这里制造的形状的体积差不多与我们刚刚计算过的饼干所用面团的体积相似（或多或少有一点儿曲度上的差异）。事实上，如果我们想让这个图形更像堆叠起来的饼干，我们应该旋转下面这些矩形条柱（见图16–24）。

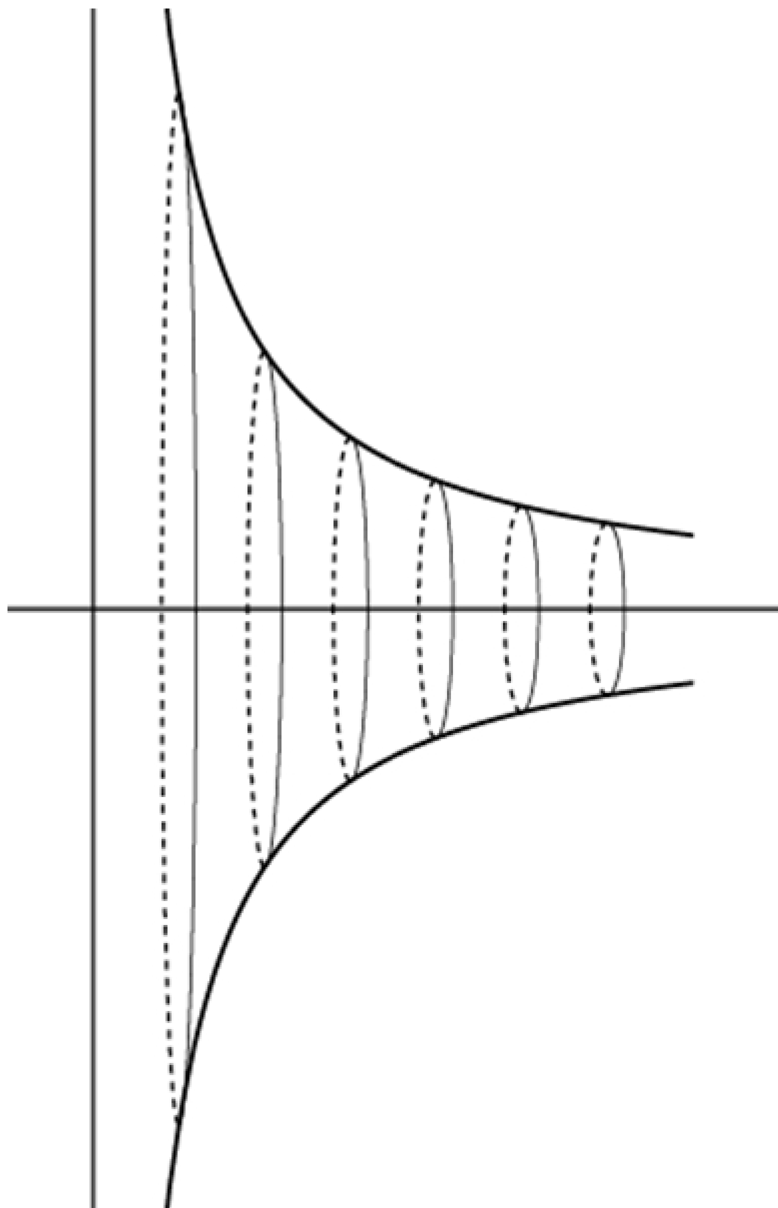


图16–23

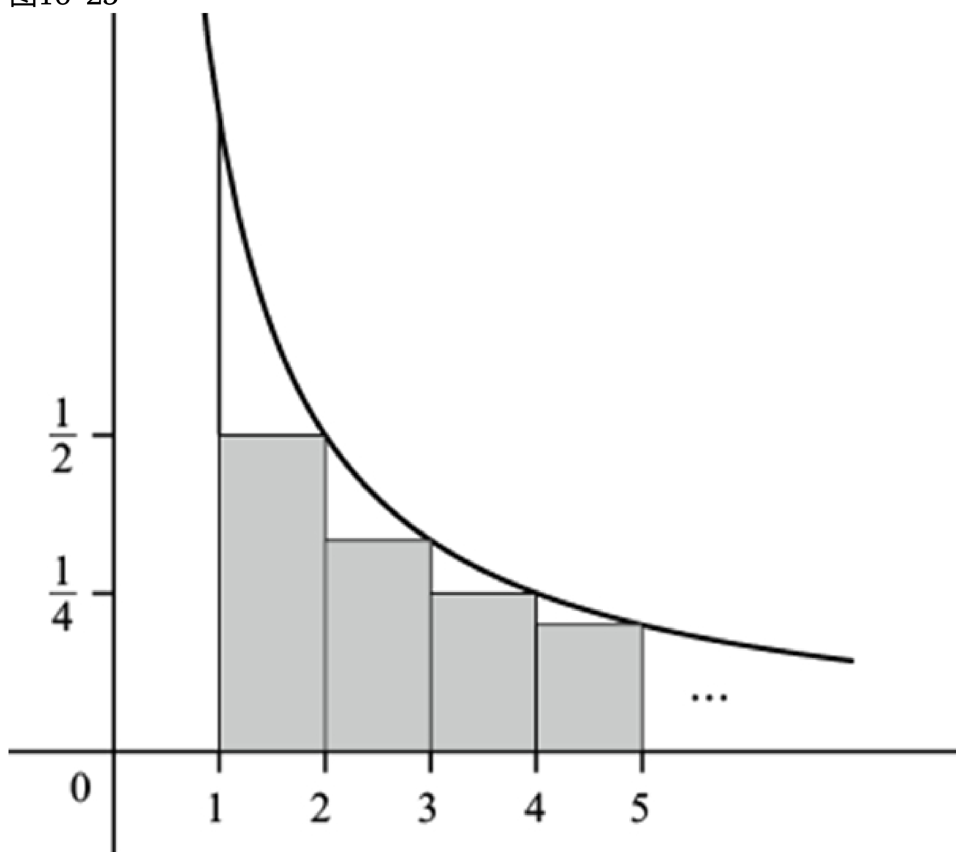


图16–24

这样我们就能够获得图16–25所示的形状。

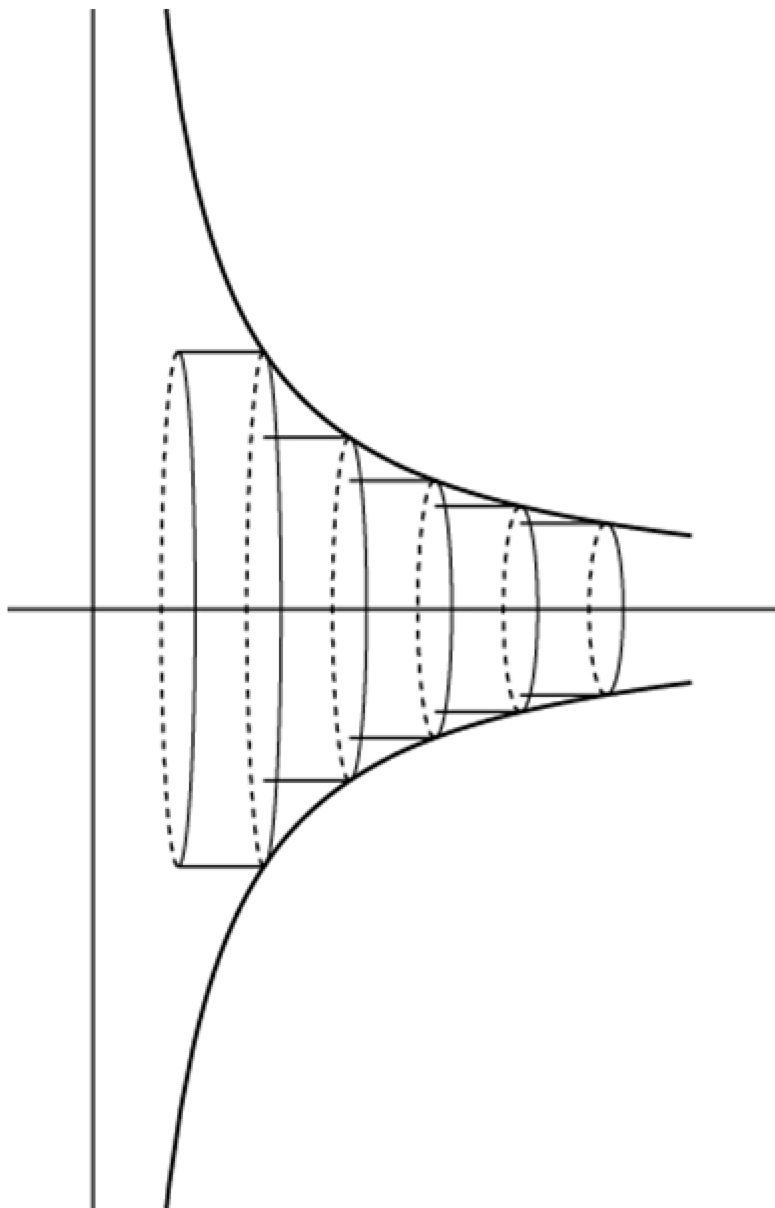


图16–25

现在，奇怪的事情出现了。矩形条柱的面积是无穷的（我们在前面已经证明过这个结论了），但是饼干堆的体积是有限的。无穷饼干的堆叠高度是无穷的，而饼干堆的横截面也是无穷的！（因为该面积是矩形条柱面积的

两倍。）

现在你开始感到不安了吧，就像坐在一条在浩瀚而狂野的海洋中上下颠簸的小船上一样。但这就是无穷的乐趣所在。对我来说，真正的乐趣是理解无穷背后的逻辑，然后观察如何用严谨的数学方法来解决这些问题。这样接下来你就可以驾驭这些波浪而不会晕船了，甚至你可能会为此感觉非常兴奋。

最近，我去悉尼参加了一个会议。在会议开始前，我参加了一个观赏鲸鱼的乘船旅行。大海波涛汹涌，但是我一直提醒自己顺应船的起起伏伏而不是试图抵抗它。在我们已经驶离港口很久之后，船上的大部分其他人都难受得不得不躺下，只有我和少数几个人能够真正欣赏座头鲸令人难以置信的力量、威严和优雅。在数学中，我们也能够获得同样的感受。数学是我们努力寻找的另外一种与众不同的力量、威严和优雅。但是遗憾的是，太多人在旅途中“晕船”了。

17　跨越逻辑边界

在我最喜欢的《小熊维尼》的故事中，有一段讲的是维尼和小猪绕着一棵树追踪一些脚印。它们认为它们正在追踪一头大象。当再次回到出发的地方时，它们并没有意识到它们在雪地里看见的脚印其实是它们自己的。所以它们认为现在有两头以上的动物加入了第一头“大象”的行列（而实际上第一头“大象”的脚印也是维尼自己的）。它们又绕着树走了好几圈才意识到事实上发生了什么，于是它们羞怯地回家了。

我曾经在本书开始的时候提到过，我小时候生活的房子中间有一个火炉，火炉上有一个烟囱（我们确实拥有一个中央加热系统），而房子里所有的房间都是围绕着火炉排布的。这意味着每一个房间都是围绕着火炉彼此相连的，而不是像很多其他房子一样，房间和入口是相连的。

当我还小的时候，这种建筑结构带给我的最好的记忆就是我和我妹妹可以在屋子里转圈追逐。我们围着火炉一圈又一圈地跑，穿过厨房、餐厅、客厅、厨房、餐厅、客厅。当然，我们中的一个人会突然改变方向，试图从背后抓住另一个。然后，我们会尖叫着朝相反的方向跑，厨房、客厅、餐厅、厨房、客厅、餐厅。这种具有类似圆形循环结构的房子的奇妙之处在于，它让你拥有了一个仿佛是无穷大的房子。也就是说，有一条无穷长的路可以供你行走。是的，你将会持续回到同一个地点。但是环绕一周后回到起点和从A走到B再回到A是有所不同的。如果你像忒修斯一样，在走迷宫时身后拖着一根绳子，那么你就会发现，围绕烟囱走一圈回到起点与从A点走到B点再返回A点是完全不同的。因为绳子发生了变化，它现在沿着烟囱绕了一圈。你经历了一个真实的旅程。

数学家们用一种被称为“覆叠空间”的概念来处理这种现象。覆叠空间的要点是，你所画的建筑物地图呈现的不再是每一个地方在哪，而是所有你能选择的路径，就像是身后拖着一根绳子的忒修斯站在迷宫里一样。想象一下，你将绳子的一端系在入口处，然后将绳子的另一端系在自己身上。那么当你进入迷宫以后你经过的所有地点都将连在一起。

如果你沿着原路返回离开迷宫的话，那么你的绳子就会随着你一起回来。但是如果你选择了一条不同的路径离开迷宫的话，那么绳子就可能会卡在迷宫里，从入口连接到出口。如果我回到儿时的房子将绳子系在前门上，然后绕着烟囱一圈又一圈地跑，那么我将需要一条无穷长的绳子（可能还需要有一个保镖防止我被绳子套住捆起来）。在覆叠空间的解释里，这种情况会被视作一条真正无穷长的路径，而不是重复绕回起点。就像小猪和维尼一样，你并不会意识到自己每次都会回到开始的地方。

我的心愿之一就是再次住在一个带有真正圆形回路的房子里。在数学家中间，我把这种房子称为“具有同伦性”的房子。因为同伦是研究空间中封闭路径的数学理论。（我们曾经在第13章提到过同伦这个概念。）如果你能够拉着忒修斯式的绳子回到起点而不会把绳子绕在某个类似于烟囱这类物体上，那么这个回路就不是一个真正的回路。只有你的绳子会绕在某个物体上的时候，这种回路才能被称为真正的回路。比如我做完膝盖手术后能走的最长的距离就是围着我的厨房悲惨而毫无目的地绕小圈，这类回路就不是真正的回路。

我对建有一个以上楼梯的房子格外有兴趣。除了这类房子通常非常宏伟以外，我还喜欢这些楼梯所提供的垂直回路。你可以从一个楼梯上去，然后从另一个楼梯下来，再回到第一个楼梯的底部。在这种情况下，你就没办法把你的忒修斯绳子收起来了。

我现在居住的地方没有真正的闭合回路，因此它没有同伦性。

无穷的路径

即使我不再经常一圈圈地追逐我妹妹，我还是很喜欢包含一些回路的建筑。因为这意味着我可以进行一场无穷长的散步。数学家们通常喜欢用散步来帮助他们思考。毫无疑问，其他很多人也是如此。这种温和的运动可以帮助我拓展思维。有时候，我认为行走这个动作能够使我大脑的后勤部分保持足够繁忙，这样我有幻想力的数学大脑就能自由漫步而不再受其他部分的无休止的打扰，比如，大脑的后勤部分就不会再说像是“不要忘记买鸡蛋”这样的事。

然而，进行一次无穷长距离的散步会带来一个问题，就是结束的时候你会发现自己距离家已经无穷远了。此外，如果我在户外散步的同时思考数学，我很可能会迷路甚至被车撞。

如果房子里自带一个回路，这将解决所有这类问题，因为这意味着你可以“永远”一圈一圈地走下去。你不会感觉这些圆圈是无意义的，因为不像我手术后只能围着厨房绕圈子，你可以绕着某个实心的事物而不是一个空荡荡的空间转圈。就像我曾经提到过的，尼斯大学的数学系系馆是一幢环形建筑。环形建筑所包围的区域是一个非常漂亮的圆形内置庭院，人们可以站在走廊的落地窗前向下看，观赏庭院。所有的办公室都在环形的外圈。所以当你沿着走廊一圈一圈走的时候，你总能面朝着这个美丽的庭院。我很高兴地发现，我不是唯一一个喜欢一边在环形走廊中散步一边思考的人。

但数学系系馆这种布局的问题就是它太对称了，因此我无法记住我的办公

室的位置。整幢楼里一共有4个楼梯，平均分布在整个圆周上。无论我什么时候从楼梯上来，我都无法判断应该往左走还是往右走才能更快到达我想去的地方。可笑的是，为了到达我想去的地方，我绕着环形一直走的圈数常常会大于一，因为我在路上错过了我的目的地。

帕特诺斯特（Paternoster）

另一个让人满意的无穷结构是帕特诺斯特。在谢菲尔德大学的艺术塔里，就有一个非常有名的帕特诺斯特结构。它像一个电梯，但是比电梯有趣得多。它由一圈循环的电梯舱组成，每一个电梯舱可以搭载两个人。任何时刻，每一层都有一个电梯舱要上行，一个电梯舱要下行。并且在循环的过程中，有一些电梯舱会超过顶层一点儿或者低于底层一点儿。帕特诺斯特电梯没有可关闭的门，而且它不会停止运动。当电梯舱经过你所站的楼层时，你要跳进电梯，再在到达你要去的那层时跳出电梯。在实际使用中，电梯移动得非常慢，但是我在最初几次搭乘电梯时还是感觉很恐怖。当我跳上和跳下电梯的时候，我都会心跳加速。令人惊奇的是，这部电梯并不违反建筑安全条例。原则上讲，学校禁止人们乘坐电梯超过顶层或者低于底层，但我猜学校里没有多少人不想尝试一下这件事。

这种结构被称为帕特诺斯特，其与念珠的构造密切相关。念珠存在的主要目的是记录你说某个祷告的次数。在一串念珠中，十个小念珠排在一起是为了让你数清你已经说了十遍万福玛利亚，这之后会有一个大念珠，它是为了提示你在说下十个万福玛利亚之前应该先说一遍主祷文。“帕特诺斯特”就是拉丁语主祷文的起始词。念珠的设计初衷是让人拥有一个物理的方法来记录自己重复的次数，这样你的大脑就有多余的空间去思考了。

我对帕特诺斯特电梯的感觉与念珠类似。我希望建筑里总是有帕特诺斯特电梯而不是普通电梯。对我而言，帕特诺斯特电梯会让人感觉整个建筑物是在一个无穷环路上相连的，而不是分层建造的，在后一种情况中，人们只能在有限尺寸的楼层间移动。这意味着当我乘坐帕特诺斯特电梯在楼层间移动时，我的大脑不会感到整段路被分成了许多层。比起等待一个不可预测的电梯的到来，可预测的持续运动意味着我的头脑有更多的空间可以自由地思考。这种说法有一点儿夸张。但是当我思考数学时，这类事物确实给我带来了大量的灵感。

除此之外，帕特诺斯特电梯还是那个“能够一圈一圈循环的事物”的垂直版本（虽然你无法真的在帕特诺斯特电梯上绕圈追逐某人）。而且帕特诺斯特电梯给我一种建筑物无穷大的感觉。如果你将你的忒修斯式绳子的一端系在第一层，然后进入帕特诺斯特电梯并随着它一圈又一圈地移动，那么你同样需要一根无穷长的绳子。

其他的圆圈

我们之前曾说过，世界上没有真正无穷的事物。但是现在我们发现，如果考虑你在圆圈上能行走的路径，那么圆圈就是一种无穷。圆形或椭圆形赛道是非常聪明的设计，因为你可以在上面进行任何长度的比赛。最近有一个叫雅各布的小男孩写信给我，他说数学是他最喜欢的学科。而且他还在信的末尾顺便告诉我，他能够在攀爬架上前进84米。我想知道，如果他有一个非常长的或者持续转动的或者环形的攀爬架，那么他是不是能一直攀爬，想爬多远就爬多远？

地铁的环形线路更有趣，因为你可以永远坐在地铁上，一圈又一圈地前进。与在一个非环形线路的两个终点站之间来回往返相比，乘坐环形线路更像是一场无穷的旅程。有些人喜欢沿着伦敦M25高速公路一圈又一圈地开车，他们为此感到很兴奋。（虽然就我个人而言，沿着M25高速公路行驶时能不完全堵在路上我就很高兴了。）我相信其他城市的环形公路上也有这样的狂热分子，比如巴黎的环城大道或者华盛顿哥伦比亚特区的首都环线。

圆圈还有一个更有趣的版本，就是莫比乌斯带。它是用一个纸条做成的，你将纸的两端粘在一起。但是并不是用下面这种平淡无奇的方式（见图17–1）：



图17–1

而是在把两端粘在一起之前翻转一下纸条的其中一端（见图17–2）。



图17–2

关于莫比乌斯带有一个有趣的事实：无论是在物理上讲还是在数学上讲，由于你已经把纸条一端的前面粘在了另一端的后面，把这一端的后面粘在了另一端的前面，这意味着这个纸条现在就只有一个面了——里面和外面变成了同一个面。

你可以用莫比乌斯带做许多有趣的事，最简单的一个就是，你可以拿起它，然后让你的手指沿着纸条移动。相比于沿着一个圆圈移动，沿着莫比乌斯带移动更容易让人迷失方向，因为你会困惑于你的手指现在是在里面还是在外面。因为里面和外面并没有区别。你还可以尝试沿着莫比乌斯带的边缘追踪你的手指。它的边缘是个简单的圆圈，但是这个圆圈在回到起点后又会“绕着”莫比乌斯带走一遍。虽然事实上它只有一圈。边缘的轨迹是一种像“8”的图形，这又是另一个可以永远一圈又一圈绕行下去的令人满意的形状。用这个图形作为我们探索无穷的起点以及终点是多么合适啊！

0　∞

但是我们知道，一切并没有结束。即使我们一直静静地坐在我们大而有限的世界里，无穷还是会持续到永远，而且无穷的等级也会持续增加到永远。也许我们已经不再惊讶于一些无穷的事物能够装载在一些有限的事物里。涉及无穷的事物，似乎一切都是可能的。数学世界也装在我们的大脑里，但它比宇宙还要大。

无穷是自由的，但有时它过于自由了。如果我可以掌控无穷的时间的话，我可以进行更加自由和有创造性的思考。如果一天里剩余的时间我没有固定安排的话，我更有可能完成一个定理的证明，即使证明这个定理只会花费两个小时。但是如果实际上我只有两个小时的时间的话，那我可能就无法证明出这个定理了。

然而我们已经发现，如果我们是不死的，那么我们就可能永远拖延着不去

工作。以我对自己的了解，我可能真的会永远拖着。生活在无穷的维度里为我们带来了无穷的我们无法解释的精妙之处。我们仍然可以幻想这些精妙之处，即使我们无法解释它们。

我真的认为试图解释清楚所有事物并不是关键。相反，我认为关键是我们尽可能多地解释我们能够解释的事物。更重要的是，我们应该清楚我们能够解释和不能够解释的边界在哪里。在我的脑海中，我们在逻辑上能够解释的事物处在观念宇宙的中心位置，而数学的目标就是将尽可能多的事物纳入这个范围。所以这个范围往往会变得越来越大。而随着范围的扩大，它的表面也会持续增长。这个表面就是我们能解释的事物和不能解释的事物的边界。

对我而言，最美丽的事物就是刚刚跨过逻辑边界的事物。我们需要走很长一段路来解释这些事物，但是最终它们还是避开了我们。我需要走很长的一段路来解释为什么一段音乐会让我哭泣。但分析到某一点之后，总会剩下一些事是我的分析无法解释的。另一个类似的问题是，为什么看着大海会让我那么兴奋，或者为什么爱情是那么美好，或者为什么无穷是那么迷人。还有些事物是我们甚至无法试图去解释的，这些事物存在于远离我们观念宇宙逻辑中心的领域里。但是对我来说，所有的美丽就刚刚好在那个边界上。随着逻辑领域内部和外部之间的分界面的扩大，我们实际上接触到了越来越多的美丽。对我来说，这就是一切。

生活中和数学中存在一个美丽与实际的权衡问题。除此之外，还有梦想与现实之间、可解释与不可解释之间的矛盾。无穷是一个美丽的梦想，这个美丽梦想的核心就是数学。

致谢

我第一个想感谢的人是我已逝去的学生丽莎·奎魏妮恩。她向我展示了芝诺悖论在漫画中的一个特别有趣的应用。我想要感谢我所有的学生，无论他们来自芝加哥美术学院，还是芝加哥大学、谢菲尔德大学或者剑桥大学。同样感谢剑桥的派克街小学、谢菲尔德的希尔斯伯勒小学、芝加哥的弗朗西斯帕克中学的孩子们。我相信教学是一个双向的过程，而我在这些年中也从我的学生们身上学到了很多东西。

如果没有我的父母、我的妹妹、我的小外甥利亚姆和杰克的支持的话，这本书不可能问世。

我还要感谢我的朋友们，他们的见解常常启发我从不同的角度思考事物，让我把事物思考得更加清晰。我要特别感谢下面这些人，我在本文中特别提到过他们的见解。阿玛妮亚·家本桃修、杰森·格鲁内鲍姆、克里斯托弗·丹尼尔森、理查德·伍德、汤姆·克劳福德、萨利·兰达尔、戴维·哈钦斯、山姆·杜普莱西斯、凯瑟琳·芬奇、艾丽丝·谢乌、杰拉尔德·芬利。感谢考特尼·兹瑞博、蒂莫西·迈登、洛翰·周–李为我提供跨文化参考。

第12章献给格雷戈瑞·皮布尔斯，向他一直以来对于维度的热爱致敬。

我想要感谢我的英语老师玛瑞斯·拉金。我从她那里学习到的文章写作方式非常适合书籍的写作。感谢我的经纪人戴安娜·班克斯。感谢Profile出版社的尼克·希兰和安德鲁·富兰克林，基本图书公司（Basic Books）的凯莱赫和劳拉·海默特。感谢莎拉·加布里埃尔，你一直是引领我走出迷雾的灯塔。感谢鲁本·托马斯为我提供Ubuntu系统下LaTeX的解决方案。感谢奥利弗·卡马乔一直以来不离不弃的陪伴。

最后，感谢阿康托旅馆的米迦勒为我提供食物、饮品，并且和我讨论芝诺。